

University of Khemis Miliana
Faculty of Matter Sciences and Computer Science
Department of Artificial Intelligence

Urban Building Footprint Extraction using Graph Neural Networks and Assessed OpenStreetMap Data with Sentinel-2 Imagery

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of
Master in Artificial Intelligence and Big Data

Prepared by:
Anouar ADEL & Maamar ARBOUZ

Supervised by:
Dr. Meziane IFTENE
Dr. Mohammed El Amin LARABI
Algerian Space Agency

Co-supervised by:
Dr. Mohamed GOUDJIL
University of Khemis Miliana

Jury Members:
President: **Dr. Abdelhamid HARICHE**
Examiner: **Dr. Imène AIT ABDERRAHIM**

Academic Year: 2024-2025

Abstract

The accurate and timely identification of urban building footprints is critical for sustainable urban planning, disaster management, and monitoring dynamic environmental changes, particularly in rapidly urbanizing regions like Algiers, Algeria. Traditional methods for footprint extraction often face limitations in scalability, accuracy, and adaptability to complex urban morphologies. This thesis addresses these challenges by developing and evaluating a novel framework that synergistically integrates Graph Neural Networks (GNNs) with superpixel segmentation of Sentinel-2 multispectral imagery.

The primary objective was to assess the efficacy of a specifically configured UrbanGraphSAGE model for semantic segmentation of building footprints. A cornerstone of the methodology involved the rigorous assessment of OpenStreetMap (OSM) data, cross-validated with Google Open Buildings and Overture Maps datasets and refined through temporal NDVI stability analysis, to generate high-quality ground truth labels. Furthermore, the impact of image-level spectral data augmentation on model performance and generalization was investigated. The proposed UrbanGraphSAGE model's performance was systematically compared against other GNN architectures, namely Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), under identical data and training conditions.

The research demonstrated the augmented UrbanGraphSAGE model's robust performance, yielding a test F1-Score of 0.7579, with a high recall of 0.9192, effectively identifying building footprints in the complex Algiers study area. The meticulous OSM data assessment strategy proved crucial for enhancing ground truth reliability, and spectral data augmentation contributed to a quantifiable improvement in model generalization (F1-Score increase of 0.0282 compared to the non-augmented baseline). Comparative analysis showed UrbanGraphSAGE and GCN performing comparably well and significantly outperforming GAT. Permutation feature importance analysis (on a non-augmented model) highlighted Short-Wave Infrared (SWIR) and Red Edge bands from Sentinel-2 as particularly discriminative.

This study validates a comprehensive and replicable framework for leveraging GNNs and assessed open data sources for detailed urban feature extraction from medium-resolution satellite imagery. The findings underscore the importance of data quality and augmentation in developing robust GeoAI solutions and provide practical insights for urban monitoring applications in dynamically evolving cities.

The developed methodology offers a significant step towards more automated, accurate, and scalable urban mapping.

Keywords: Urban Building Footprint Extraction, Graph Neural Networks (GNNs), Sentinel-2 Imagery, OpenStreetMap (OSM), Superpixel Segmentation, Deep Learning, GeoAI.

Résumé

L'identification précise et rapide des empreintes de bâtiments urbains est essentielle pour la planification urbaine durable, la gestion des catastrophes et le suivi des changements environnementaux dynamiques, en particulier dans des régions à urbanisation rapide comme Alger, Algérie. Les méthodes traditionnelles d'extraction d'empreintes se heurtent souvent à des limites en termes de scalabilité, de précision et d'adaptabilité aux morphologies urbaines complexes. Cette thèse aborde ces défis en développant et en évaluant un cadre novateur qui intègre de manière synergique les Réseaux de Neurones sur Graphes (GNNs) avec la segmentation en superpixels de l'imagerie multispectrale Sentinel-2.

L'objectif principal était d'évaluer l'efficacité d'un modèle UrbanGraphSAGE spécifiquement configuré pour la segmentation sémantique des empreintes de bâtiments. Une pierre angulaire de la méthodologie a consisté en l'évaluation rigoureuse des données OpenStreetMap (OSM), validées par croisement avec les jeux de données Google Open Buildings et Overture Maps, et affinées par une analyse de la stabilité temporelle du NDVI, afin de générer des étiquettes de vérité terrain de haute qualité. De plus, l'impact de l'augmentation des données spectrales au niveau de l'image sur les performances et la généralisation du modèle a été étudié. Les performances du modèle UrbanGraphSAGE proposé ont été systématiquement comparées à celles d'autres architectures GNN, à savoir les Réseaux Convolutionnels sur Graphes (GCN) et les Réseaux à Attention sur Graphes (GAT), dans des conditions de données et d'entraînement identiques.

La recherche a démontré la performance robuste du modèle UrbanGraphSAGE augmenté, obtenant un F1-Score de test de 0.7579, avec un rappel élevé de 0.9192, identifiant efficacement les empreintes de bâtiments dans la zone d'étude complexe d'Alger. La stratégie méticuleuse d'évaluation des données OSM s'est avérée cruciale pour améliorer la fiabilité de la vérité terrain, et l'augmentation des données spectrales a contribué à une amélioration quantifiable de la généralisation du modèle (augmentation du F1-Score de 0.0282 par rapport à la référence non augmentée). L'analyse comparative a montré que UrbanGraphSAGE et GCN ont des performances comparables et surpassent de manière significative GAT. L'analyse de l'importance des caractéristiques par permutation (sur un modèle non augmenté) a mis en évidence les bandes Infrarouge à Courtes Longueurs d'Onde (SWIR) et Red Edge de Sentinel-2 comme étant particulièrement discriminantes.

Cette étude valide un cadre complet et reproductible pour exploiter les GNNs et les sources de données ouvertes évaluées pour l'extraction détaillée de caractéristiques urbaines à partir d'imagerie satellitaire à moyenne résolution. Les résultats soulignent l'importance de la qualité des données et de l'augmentation dans le développement de solutions GeoAI robustes et fournissent des perspectives pratiques pour les applications de surveillance urbaine dans des villes en évolution dynamique. La méthodologie développée offre une avancée significative vers une cartographie urbaine plus automatisée, précise et scalable.

Mots-clés : Extraction d'Empreintes de Bâtiments Urbains, Réseaux de Neurones sur Graphes (GNNs), Imagerie Sentinel-2, OpenStreetMap (OSM), Segmentation en Superpixels, Apprentissage Profond, GéoIA.

ملخص

يُعد التحديد الدقيق والسرع ب بصمات المباني الحضرية أمراً بالغ الأهمية للتخطيط الحضري المستدام، وإدارة الكوارث، ورصد التغيرات البيئية الديناميكية، لا سيما في المناطق سريعة التحضر مثل الجزر العاشرة. غالباً ما تواجه الطرق التقليدية لاستخلاص بصمات المباني قيوداً في قابلية التوسيع والدقة والقدرة على التكيف مع الأشكال الحضرية المعقدة. تتصدى هذه الأطروحة لهذه التحديات من خلال تطوير وتقييم إطار عمل مبتكر يدمج بشكل تآزري الشبكات العصبية على الرسوم البيانية (GNNs) مع تجزئة الصور الفائقة (superpixel segmentation) لصور Sentinel-2 متعددة الأطياف.

كان الهدف الأساسي هو تقييم فعالية نموذج UrbanGraphSAGE المصمم خصيصاً للتجزئة الدلالية لبصمات المباني. تمثل حجر الزاوية في المنهجية التقييم الدقيق لبيانات OpenStreetMap (OSM)، مع التحقق من صحتها بمقارنتها بجموعات بيانات Google Open Buildings و Over-ture Maps، وتحسينها من خلال تحليل استقرار مؤشر NDVI الزمني، لتوليد بيانات حقيقة ميدانية عالية الجودة. علاوة على ذلك، تم بحث تأثير زيادة البيانات الطيفية على مستوى الصورة على أداء النموذج وقدرته على التعميم. تمت مقارنة أداء نموذج UrbanGraphSAGE المقترن بشكل منهجي مع معماريات GNN أخرى، وهي الشبكات التلافيفية على الرسوم البيانية (GCN) وشبكات الانتباه على الرسوم البيانية (GAT)، تحت نفس ظروف البيانات والتدريب.

أظهر البحث الأداء القوي لنموذج UrbanGraphSAGE المعزز، حيث حقق F1-Score في الاختبار بقيمة 0.7579، مع استدعاء عالٍ قدره 0.9192، مما يحدد بصمات المباني بفعالية في منطقة الدراسة المعقدة بالجزائر العاصمة. أثبتت استراتيجية تقييم بيانات OSM الدقة أنها حاسمة في تعزيز موثوقية الحقيقة الميدانية، وساهمت زيادة البيانات الطيفية في تحسين قابل للقياس في تعميم النموذج (زيادة F1-Score بمقدار 0.0282 مقارنة بالنموذج الأساسي غير المعزز). أظهر التحليل المقارن أن تحويل أهمية الخصائص بالتبديل (على نموذج غير معزز) الضوء على نطاقات الأشعة تحت الحمراء قصيرة الموجة (SWIR) و Red Edge من Sentinel-2 باعتبارها الأكثر تميزاً.

تؤكد هذه الدراسة صحة إطار عمل شامل وقابل للتكرار للاستفادة من GNNs ومصادر البيانات المفتوحة المقيدة لاستخلاص الميزات الحضرية التفصيلية من صور الأقمار الصناعية متوسطة الدقة. تؤكد النتائج على أهمية جودة البيانات وزيادتها في تطوير حلول ذكاء اصطناعي جغرافي متينة، وتقدم رؤى عملية لتطبيقات الرصد الحضري في المدن المتطرفة ديناميكياً. تقدم المنهجية المطورة خطوة مهمة نحو رسم خرائط حضرية أكثر أمنة ودقة وقابلية للتتوسيع.

الكلمات المفتاحية: استخلاص بصمات المباني الحضرية، الشبكات العصبية على الرسوم البيانية (GNNs)، صور Sentinel-2، OpenStreetMap (OSM)، تجزئة الصور الفائقة، التعلم العميق، الذكاء الاصطناعي الجغرافي (GeoAI).

Acknowledgements

First and foremost, we would like to express our deepest gratitude to our supervisors: Dr. MezianeIFTENE and Mr. Mohammed El Amin LARABI from the Algerian Space Agency (ASAL), and Dr. Mohamed GOUDJIL from the University of Khemis Miliana. Their invaluable guidance, unwavering support, and insightful feedback were instrumental throughout this research journey. Their collective expertise not only shaped this thesis but also profoundly enhanced our understanding of the field.

We extend our sincere appreciation to the Algerian Space Agency for providing the institutional support and resources that made this research possible. The opportunity to work within such a dynamic and forward-thinking organization has been a great privilege.

We are also grateful to the Faculty of Matter Sciences and Computer Science at the University of Khemis Miliana for providing the academic framework for this Master's program.

Our thanks go out to our colleagues and friends who offered encouragement and engaged in stimulating discussions that helped refine our ideas.

Finally, we owe a special debt of gratitude to our families for their endless love, patience, and support. This accomplishment would not have been possible without them.

Contents

Abstract	1
Résumé	1
1	ملخص
Acknowledgements	1
Contents	2
List of Figures	6
List of Figures	9
List of Tables	10
List of Tables	10
List of Abbreviations	11
1 Introduction	13
1.1 Background and Motivation	13
1.2 Remote Sensing for Monitoring Building Footprints	15
1.3 Deep Learning and Graph Neural Networks for Building Footprint Extraction	16
1.4 The Role of Superpixel Segmentation	18
1.5 Leveraging and Assessing OpenStreetMap Data for Reliable Ground Truth	20
1.6 Research Questions and Objectives	21
1.7 Scope and Limitations	23
1.8 Thesis Structure	25

2 Literature Review and Theoretical Framework	27
2.1 Introduction	27
2.2 Remote Sensing for Building Footprint Extraction	28
2.2.1 Principles of Satellite-Based Building Footprint Identification	28
2.2.2 Key Characteristics of Sentinel-2 Relevant to Building Footprint Extraction	30
2.2.3 Commonly Used Spectral Indices and Their Relevance to Building Identification	32
2.2.4 Essential Preprocessing of Sentinel-2 Data for Urban Building Footprint Extraction	34
2.3 Artificial Intelligence in Remote Sensing for Feature Extraction .	35
2.3.1 The Paradigm Shift: AI, ML, and DL in Earth Observation	35
2.3.2 Traditional Machine Learning Approaches for Building Footprint Mapping and Their Limitations	36
2.3.3 Deep Learning and CNNs for Semantic Segmentation of Building Footprints	38
2.4 Superpixel Segmentation and Graph Neural Networks (GNNs) .	40
2.4.1 Superpixel Segmentation: Concepts and Relevance to Graph-Based Image Analysis	40
2.4.2 Graph Neural Networks (GNNs): Theoretical Foundations and Suitability for Graph-Structured Data	41
2.4.3 Synergistic Application of Superpixels and GNNs for Building Footprint Extraction	43
2.5 OpenStreetMap (OSM) Data for Ground Truth in Earth Observation	45
2.5.1 Characteristics of OSM Data: VGI, Building Tags, and Open Licensing	45
2.5.2 Known Data Quality Considerations for OSM Building Footprints	46
2.5.3 Importance and Methods of OSM Data Quality Assessment for Building Footprints	48
2.5.4 Utilizing Google Open Buildings and Overture Maps for OSM Cross-Validation	49
2.5.5 Critical Role of OSM Data Assessment for GNN Model Reliability	51
2.6 Conclusion	52
3 Methodology	55
3.1 Problem Formulation	55
3.2 Data Acquisition and Preprocessing	58

3.2.1	Study Area	58
3.2.2	Satellite Imagery Acquisition and Preprocessing	59
3.2.3	Ground Truth Data Acquisition and Assessment Strategy	63
3.3	Superpixel Segmentation and Graph Construction	72
3.3.1	Superpixel Segmentation using SLIC	73
3.3.2	Graph Construction from Superpixels	75
3.4	AI/Deep Learning Model Development: UrbanGraphSAGE	77
3.4.1	Model Architecture: UrbanGraphSAGE	78
3.4.2	Training Methodology for UrbanGraphSAGE	82
3.4.3	Feature Importance and Selection Considerations	83
3.5	Implementation and Evaluation	85
3.5.1	Evaluation Metrics	85
3.5.2	Evaluation Protocol	87
3.5.3	Baselines for Comparison	88
3.5.4	Software and Hardware Environment	88
4	Results and Discussion	90
4.1	Presentation of Results	90
4.1.1	Performance of UrbanGraphSAGE Model (with Spectral Augmentation and Assessed OSM Data)	90
4.1.2	Comparative Performance with GCN and GAT Models	93
4.1.3	Impact of Spectral Data Augmentation	96
4.1.4	Insights from SLIC Parameter Experimentation	97
4.1.5	Findings from Permutation Feature Importance (on Non-Augmented Model)	100
4.1.6	Qualitative Results: Visual Assessment of Building Footprint Extraction	103
4.2	Analysis and Interpretation of Results	108
4.2.1	Performance Analysis of the Augmented UrbanGraphSAGE Model	109
4.2.2	Impact of Assessed OSM Data and Spectral Augmentation on Model Robustness	110
4.2.3	Comparative Analysis of GNN Architectures (UrbanGraphSAGE vs. GCN vs. GAT)	111
4.2.4	Interpretation of Feature Importance and Selection (Non-Augmented Model Context)	112
4.2.5	Observed Challenges and Limitations in Building Footprint Extraction	113
4.3	Discussion (Contextualization and Broader Implications)	114

4.3.1	Answering Research Questions	115
4.3.2	Contribution to the Field	116
4.3.3	Implications for Urban Planning and Management in Algiers	117
4.3.4	Limitations of the Study	118
4.3.5	Recommendations for Future Work	119
5	Conclusion	121
5.1	Restatement of Research Objectives and Scope	121
5.2	Summary of Key Findings	122
5.3	Significance and Contributions of the Research	124
5.4	Limitations of the Study	126
5.5	Concluding Remarks	127
Bibliography		128
A Profile of the Algerian Space Agency (ASAL)		135
A.1	Introduction	135
A.2	Presentation of the Host Organization: ASAL	135
A.2.1	Historical Overview	135
A.2.2	Organizational Structure	136
A.2.3	Logo of the Algerian Space Agency (ASAL)	137
A.3	ASAL's Operational Domains and Strategic Engagements	137
A.3.1	ASAL Core Missions	137
A.3.2	Satellite Infrastructure and Programs	138
A.3.3	Remote Sensing and Technological Applications	138
A.3.4	International Collaboration and Research Opportunities .	139
A.4	Conclusion	139

List of Figures

3.1	Map illustrating the selected Area of Interest (AOI) in Algiers, Algeria, used for this study. The rectangular AOI is shown by the blue overlay. The boundary coordinates defining this AOI are provided in Section 3.2.1. (Base map data source: OpenStreetMap contributors)	59
3.2	Conceptual diagram illustrating the key steps in the Sentinel-2 data preprocessing pipeline, from raw Level-2A product acquisition to the final 15-channel normalized composite image ready for super-pixel segmentation. (Source: Author)	62
3.3	OpenStreetMap (OSM) building footprints retrieved for the defined Area of Interest (ROI) in Algiers prior to quality assessment and cross-validation. The ROI boundary is shown in blue. The plot shows 75921 OSM building features within the AOI. (Source: OpenStreetMap contributors, data fetched via Overpass API and plotted by author)	64
3.4	Conceptual flowchart illustrating the multi-source cross-validation process for OpenStreetMap (OSM) data quality assessment. The process involves spatial joins between OSM, Google Open Buildings (GOB), and Overture Maps building datasets, followed by confidence scoring based on inter-dataset agreement, and finally deduplication to produce a unique set of assessed building footprints. (Source: Author)	67
3.5	Conceptual diagram showing the logic for categorizing building footprints based on temporal NDVI analysis. This includes NDVI calculation from multi-date Sentinel-2 imagery, extraction of zonal statistics for Confidence Score 1 building footprints, NDVI difference calculation, and subsequent threshold-based categorization to assess temporal stability. (Source: Author)	70

3.6	Visual example of a portion of the final binary ground truth mask (<code>aligned_reliable_urban_mask.tif</code>) for the Algiers AOI. Pixels representing building footprints (value 1) are shown in white, contrasted with non-building areas (value 0) in black. This raster mask served as the definitive ground truth for labeling superpixels. (Source: Author, derived from assessed OSM, GOB, and Overture data)	72
3.7	Visual example illustrating the SLIC superpixel segmentation (boundaries shown in yellow) overlaid on a Sentinel-2 RGB composite for a portion of the Algiers study area. This demonstrates the generated superpixel shapes using the chosen parameters (requested <code>n_segments=20,000</code> , <code>compactness=1</code> , <code>sigma=1</code>), resulting in 17,916 actual superpixels for the original composite image. (Source: Author, derived from Sentinel-2 imagery)	74
3.8	Conceptual diagram illustrating the graph construction process. Stage 1 involves SLIC superpixel segmentation of the preprocessed Sentinel-2 composite image. Stage 2 details the derivation of graph components: identifying unique superpixels as nodes, extracting a 19-dimensional feature vector (15 spectral, 4 geometric) for each node, determining spatial adjacency (4-connectivity) to form the edge index, and assigning binary node labels based on majority overlap with the final ground truth building mask. Stage 3 shows the assembly of these components into the attributed superpixel graph, a PyTorch Geometric Data object ready for GNN processing. (Source: Author)	77
3.9	Architecture of the UrbanGraphSAGE model. The model takes N x 19 node features as input and processes them through three SAGE-Conv blocks (each comprising SAGEConv, BatchNorm, ReLU, and Dropout). Outputs from all SAGEConv blocks are concatenated by a JumpingKnowledge ('cat' mode) layer. The resulting N x 192 features are then passed to a final linear classifier with a Sigmoid activation function to produce N x 1 node classification probabilities. (Source: Author)	81
4.1	Visual representation of the confusion matrix for the UrbanGraphSAGE model (trained with spectral augmentation and assessed OSM data) on the test set. The matrix illustrates True Positives (TP=717), True Negatives (TN=1513), False Positives (FP=395), and False Negatives (FN=63). (Source: Author)	92

4.2	Visual representation of the confusion matrix for the UrbanGCN_Enhanced model (trained with spectral augmentation and assessed OSM data) on the test set, illustrating True Positives (TP=665), True Negatives (TN=1588), False Positives (FP=320), and False Negatives (FN=115). (Source: Author)	94
4.3	Visual representation of the confusion matrix for the UrbanGAT model (trained with spectral augmentation and assessed OSM data) on the test set, illustrating True Positives (TP=635), True Negatives (TN=1365), False Positives (FP=543), and False Negatives (FN=145). (Source: Author)	95
4.4	Comparison of Test F1-Scores achieved by UrbanGraphSAGE, UrbanGCN_Enhanced, and UrbanGAT models. All models were trained with spectral augmentation and assessed OpenStreetMap (OSM) data. (Source: Author)	95
4.5	Illustrative comparison of SLIC segmentation results on a representative zoomed-in area of Algiers, showcasing the impact of varying the <code>compactness</code> parameter. All images used requested <code>n_segments=10,000</code> and <code>sigma=1</code> . (a) Compactness=0.1 shows highly irregular superpixels strongly adhering to spectral boundaries. (b) Compactness=0.5 shows a better balance. (c) Compactness=1 (chosen for final GNN pipeline along with <code>n_segments=20,000</code>) offers a good compromise between boundary adherence and regularity. (d) Compactness=10 results in very regular, grid-like superpixels often disregarding feature boundaries. (Source: Author, derived from Sentinel-2 imagery)	100
4.6	Permutation feature importance for the non-augmented UrbanGraphSAGE model (JK='cat', 0.7F+0.3D). Importance is measured as the drop in Validation F1-Score from a baseline of 0.7350 when each feature is permuted. Features are ranked from most to least important (top to bottom). (Source: Author, generated from permutation importance analysis detailed in Section 3.4.3)	102
4.7	Example of successful building footprint extraction by the augmented UrbanGraphSAGE model in an area with high-density urban fabric. Predicted footprints (yellow overlay) are displayed on the Sentinel-2 RGB composite. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])	105

4.8 Example of successful building footprint extraction by the augmented UrbanGraphSAGE model in an area with moderate-density urban fabric and well-separated structures. Predicted footprints (yellow overlay) are displayed on the Sentinel-2 RGB composite. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])	106
4.9 Example illustrating challenges in delineating very small or fragmented building footprints. The yellow overlay shows where the model has potentially missed small structures or struggled with their precise boundaries. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])	107
4.10 Example illustrating a commission error due to spectral confusion. The model has incorrectly classified a large area of dark bare soil as 'building' (yellow overlay). (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])	108
A.1 ASAL Official Logo.	137

List of Tables

3.1	Key Characteristics of Ground Truth Data Sources Utilized	65
3.2	UrbanGraphSAGE Model Architecture Summary	79
3.2	79
3.3	Permutation Feature Importance Ranking for the Non-Augmented UrbanGraphSAGE Model (Validation F1 Baseline: 0.7350)	84
3.3	84
3.4	Evaluation Metrics Definitions	86
3.4	86
4.1	Performance Metrics of the UrbanGraphSAGE Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.	91
4.2	Confusion Matrix for the UrbanGraphSAGE Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.	91
4.3	Comparative Test F1-Scores of UrbanGraphSAGE, UrbanGCN_Enhanced, and UrbanGAT Models (All trained with spectral augmentation and assessed OSM data).	93
4.4	Confusion Matrix for the UrbanGCN_Enhanced Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.	94
4.5	Confusion Matrix for the UrbanGAT Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.	94
4.6	Permutation Feature Importance Ranking for the Non-Augmented UrbanGraphSAGE Model (Validation F1 Baseline: 0.7350). Features are ranked by the drop in Validation F1-Score when permuted.	101
4.6	101
4.7	Summary of Test Set Performance for Non-Augmented UrbanGraphSAGE Model with Reduced Feature Sets (Baseline Test F1-Score with all 19 Features: 0.7297).	103
A.1	ASAL Space Technologies Appropriation.	138

List of Abbreviations

- AI** Artificial Intelligence
- AOI** Area of Interest
- BSI** Bare Soil Index
- CNN** Convolutional Neural Network
- CRS** Coordinate Reference System
- FCN** Fully Convolutional Network
- GAT** Graph Attention Network
- GCN** Graph Convolutional Network
- GNN** Graph Neural Network
- GOB** Google Open Buildings
- GPU** Graphics Processing Unit
- IoU** Intersection over Union
- JK** JumpingKnowledge
- ML** Machine Learning
- NBI** New Built-up Index
- NDBI** Normalized Difference Built-up Index
- NDSI** Normalized Difference Soil Index
- NDVI** Normalized Difference Vegetation Index
- NIR** Near-Infrared
- OSM** OpenStreetMap

ReLU Rectified Linear Unit

RGB Red, Green, and Blue

SLIC Simple Linear Iterative Clustering

SWIR Short-Wave Infrared

UTM Universal Transverse Mercator

VGI Volunteered Geographic Information

VHR Very High Resolution

Chapter 1

Introduction

1.1 Background and Motivation

The accurate delineation of building footprints, representing the spatial extent of constructed buildings, is a critical task in an era of unprecedented global urbanization. These footprints serve as a fundamental data layer for a multitude of applications, including urban studies, infrastructure planning, population estimation, environmental monitoring, and disaster management [59]. While the broader term "urban footprint" can encompass associated infrastructure like transportation networks and other impervious surfaces, its precise definition varies. For clarity and to directly align with the primary data sources and methodologies employed, this thesis will focus on the extraction of building footprints as the core component of the urbanized landscape [48]. Understanding this targeted definition is essential, as it directly influences the data assessment strategies, the training of the proposed model, and the interpretation of subsequent results related to built-up structures.

Globally, urbanization is an accelerating trend, with more than 58% of the world's population currently residing in urban areas, a figure projected to reach 68% by the year 2050 [58]. This rapid shift towards urban living presents a multitude of challenges, including increased land consumption to accommodate new constructions, environmental degradation, strain on existing infrastructure, and the emergence of socioeconomic disparities [58]. These challenges are particularly pronounced in regions experiencing rapid development, such as Algiers, Algeria, the intended case study area for this research. Algeria faces significant pressures related to housing shortages, the proliferation of informal settlements, and growing environmental concerns. The dynamic nature of building development in such regions, often coupled with planning and implementation complexities, underscores the critical importance of developing advanced techniques for accurately mapping

and monitoring the expansion of built-up areas. This research directly supports the national strategic goals for sustainable development, a core mission of the Algerian Space Agency (ASAL), the host institution for this work. A detailed profile of ASAL’s mandate, history, and programs is provided in Appendix A.

Accurate and timely information regarding building footprints is vital for sustainable development and effective governance. Specific applications include supporting sustainable urban development initiatives by providing data on housing stock and density, monitoring environmental changes by tracking the encroachment of built structures into sensitive areas, managing risks associated with disasters by identifying vulnerable buildings [64], facilitating informed land-use planning based on existing building layouts [61], and assessing the broader impacts of urban expansion on both human and natural systems. Tools and datasets that accurately map building footprints are therefore invaluable for planners and policymakers. The ability to precisely delineate building footprints is fundamental, as inaccurate or outdated data can lead to flawed planning decisions and inefficient resource allocation [62].

Traditional approaches to mapping building footprints present several limitations. Manual digitization from aerial or satellite imagery, while often accurate, is inherently time-consuming, labor-intensive, costly, and susceptible to human interpretation errors, making it unsuitable for frequent updates over large areas [62]. Spectral index methods (e.g., NDBI, NBI, BSI), designed to highlight built-up areas based on their spectral characteristics, often struggle to accurately distinguish buildings from spectrally similar features like bare land, roads, or certain dry vegetation, leading to potential inaccuracies in building footprint delineation [56]. Furthermore, traditional machine learning approaches (e.g., SVM, Random Forest) applied to this task may require extensive feature engineering and can encounter challenges in effectively handling the complex spatial arrangements and spectral variability characteristic of dense and diverse building structures [51]. The inherent limitations in the scalability and update frequency of these conventional methods often prove insufficient to keep pace with the rapid and continuous changes occurring in built landscapes. The significant time and resource investment required for manual updates, coupled with the rapid obsolescence of the data, necessitates the exploration of more automated and advanced techniques for building footprint extraction.

1.2 Remote Sensing for Monitoring Building Footprints

Satellite remote sensing has emerged as a vital technology for acquiring comprehensive and repetitive data over vast urban expanses, offering a cost-effective and time-efficient alternative to traditional, resource-intensive ground-based surveys [51]. The ability of satellites to provide a synoptic view of urban areas, coupled with their capacity for frequent revisits, makes them indispensable tools for monitoring urban growth and change over time. In this context, the Copernicus Sentinel-2 mission stands out as a particularly well-suited source of data for such monitoring applications [52]. As part of the European Union’s Copernicus program, Sentinel-2 offers a technically feasible and sustainable solution for observing and analyzing urban environments.

The Sentinel-2 mission, comprising two twin satellites, boasts several key characteristics that make its imagery highly valuable for the task of building footprint extraction. First, its high spatial resolution, with 10 meters for the visible and near-infrared (NIR) bands (B2, B3, B4, B8) and 20 meters for several red-edge and shortwave infrared (SWIR) bands (e.g., B5, B6, B7, B8A, B11, B12) [22], allows for the detailed observation of urban structures, individual buildings, and the overall urban fabric. This level of spatial detail is crucial for accurately delineating the boundaries of built-up areas, especially when using a superpixel-based approach as adopted in this thesis. Second, the mission’s high temporal resolution, with a 5-day revisit frequency at the equator achieved by the two satellites working in concert (under cloud-free conditions) [22], enables the frequent monitoring of dynamic changes occurring within urban areas. This is particularly important in rapidly urbanizing regions like Algiers, where the extent of built-up areas can expand significantly over relatively short periods, and also supports the temporal assessment of ground truth data. Third, Sentinel-2 acquires data across a wide range of the electromagnetic spectrum, with 13 multispectral bands spanning the visible, near-infrared, and shortwave infrared regions [22]. These multispectral capabilities are essential for differentiating various land cover types present in urban landscapes, including built-up areas, vegetation, and water bodies. The specific spectral signatures captured by these bands provide valuable information for land cover classification and the identification of materials characteristic of built-up structures, forming the basis for the spectral features used in the GNN model. Fourth, the Sentinel-2 data is provided freely and openly through the Copernicus Data Space Ecosystem (formerly Copernicus Open Access Hub) [22]. This open access policy makes Sentinel-2 a highly cost-effective data source, especially when compared to commercially available Very High Resolution (VHR) imagery,

facilitating research and operational applications [52].

The combination of Sentinel-2's spatial and temporal resolution, along with its rich spectral information, facilitates both detailed and frequent monitoring of changes in built-up areas and the progress of urban development and reconstruction [54]. This is particularly crucial for capturing the dynamic nature of urban environments. The frequent revisit times allow for time-series analysis, which can provide valuable insights into expansion patterns and, as utilized in this thesis (Section 3.2.3), assess the temporal stability of features for ground truth refinement. Furthermore, specific spectral bands within the Sentinel-2 dataset exhibit sensitivity to the spectral characteristics of urban materials, making them particularly valuable for the task of built-up area identification [54]. For instance, the blue band (B02) can aid in identifying man-made features, while the red band (B04) is known to be useful for delineating built-up extents. The shortwave infrared (SWIR) bands (B11 and B12) are particularly effective in distinguishing between built-up regions and bare soil due to their sensitivity to moisture content and material composition [57]. Consequently, various spectral indices, such as the Normalized Difference Built-up Index (NDBI) and other indices like the Bare Soil Index (BSI) and the New Built-up Index (NBI), derived from combinations of these bands, have been widely utilized for mapping built-up areas and differentiating them from other land cover types [5]. The strategic selection and combination of these bands and derived indices, as detailed in the methodology (Section 3.2.2), formed the comprehensive 15-channel input for the superpixel segmentation and subsequent feature extraction stages. Combinations, such as True Color (RGB) composites and False Color Urban composites, can further enhance the visualization and interpretation of built-up extents within Sentinel-2 imagery.

1.3 Deep Learning and Graph Neural Networks for Building Footprint Extraction

Significant advancements in the fields of Artificial Intelligence (AI) and particularly deep learning techniques have brought about a revolution in remote sensing image analysis. These advanced computational methods offer automated and efficient solutions for tackling complex tasks such as land cover classification and, crucially for this thesis, the extraction of building footprints from satellite imagery [51]. Notably, AI, and specifically deep learning algorithms, have demonstrated a remarkable capability to enhance the quality of geospatial data analysis by automating various processing steps and significantly improving the accuracy of feature extraction [40]. Unlike traditional methods that often rely on manually

defined spectral and textural features, deep learning algorithms possess the inherent ability to autonomously learn and extract intricate, high-level hierarchical features directly from the raw imagery [23]. This capacity for automated feature learning is especially pivotal when dealing with the large volumes of remote sensing data acquired by missions like Sentinel-2. The ability of these algorithms to discern complex patterns and relationships within the data can lead to more accurate and robust building footprint extraction results than previously achievable with conventional techniques.

In recent years, Graph Neural Networks (GNNs) have emerged as a powerful and versatile class of deep learning models specifically designed to operate on data structured as graphs [45]. This is a significant departure from traditional neural networks, such as Convolutional Neural Networks (CNNs), which are typically applied to Euclidean data like images or sequences. GNNs are adept at handling non-Euclidean, irregular data structures. This characteristic makes them particularly well-suited for remote sensing image analysis where an image can be segmented into superpixels—perceptually meaningful regions—which are then represented as nodes in a graph. The spatial relationships (e.g., adjacency) between these superpixels define the edges of the graph [53]. In the context of image analysis, including remote sensing imagery, images can be effectively represented as graphs where individual pixels or, more commonly for efficiency and contextual representation as in this study, superpixels, serve as the nodes of the graph. The spatial relationships between these nodes, such as adjacency or proximity, define the edges [45]. GNNs are specifically architected to be directly applicable to such graph-structured data, enabling predictions at the node, edge, and even the entire graph level [45].

A key advantage of GNNs lies in their ability to capture both local and global contextual information present within the image. They achieve this by iteratively aggregating and processing information across the interconnected nodes of the graph through a mechanism known as message passing [36]. This allows the network to learn meaningful representations for each node by considering not only its own features (e.g., mean spectral values of a superpixel) but also the features of its neighboring nodes, thereby capturing the crucial contextual information inherent in spatial data [53].

The potential advantages of employing GNNs for building footprint extraction from satellite imagery, particularly when compared to more traditional Convolutional Neural Networks (CNNs), are significant [57]. First, GNNs are inherently designed to process non-Euclidean, irregular graph-structured data, such as the graphs formed by superpixels from an image segmentation. This contrasts with CNNs, which are primarily architected to operate on regular grid data formats

like standard images. CNNs often face challenges when applied directly to graph data due to the irregular shapes and varying numbers of neighboring nodes that characterize such structures [45]. Second, GNNs can excel at capturing long-range dependencies and understanding the global context within a graph, an ability that can be limited in CNNs due to their reliance on localized convolutional kernels. GNNs, through their message-passing mechanisms, can effectively propagate information across the entire graph, enabling the capture of dependencies between nodes that are far apart in the spatial domain [36]. Third, GNNs are specifically designed to leverage the relational information explicitly encoded in the graph structure, making them particularly well-suited for analyzing the spatial arrangement of built-up structures as represented by interconnected superpixels [53]. They can model the complex spatial relationships between superpixels representing potential building segments and their surrounding context, which is crucial for accurate building footprint extraction. While CNNs have achieved considerable success in various image classification tasks, GNNs offer a more natural and, as investigated in this thesis, potentially more effective approach to modeling the inherent graph structure of landscapes when represented as a network of superpixels. The graph convolution operation at the heart of GNNs has been shown to enhance feature learning specifically when dealing with graph-structured data, suggesting a potential for superior performance in building footprint extraction compared to CNN-based methods for certain problem formulations [57]. This exploration aligns with the research direction of investigating advanced AI models for Earth Observation tasks.

1.4 The Role of Superpixel Segmentation

Superpixel segmentation represents a critical preprocessing step in the analysis of remote sensing imagery, especially when preparing data for Graph Neural Network (GNN) based approaches, such as the one employed in this thesis for building footprint extraction. Superpixels are perceptually meaningful regions generated by grouping spatially contiguous pixels that exhibit similar visual characteristics, such as color, intensity, or texture [12]. This process effectively transforms the image from a granular grid of individual pixels into a more manageable set of homogeneous "atomic regions" or segments, which can then serve as the nodes in a graph representation of the image [53].

The utilization of superpixel segmentation offers several key advantages in the context of this research, particularly when interfacing with GNNs for analyzing satellite imagery:

- **Computational Efficiency:** By reducing the number of processing units from potentially millions of pixels in a satellite scene to a few thousands of superpixels, the computational complexity of subsequent graph construction and GNN operations is significantly lowered. This makes the analysis of large remote sensing datasets more tractable and efficient [54].
- **Preservation of Object Boundaries:** Well-designed superpixel algorithms, such as Simple Linear Iterative Clustering (SLIC) used in this study, tend to generate segments whose boundaries align closely with the natural contours of objects present in the image, including urban features like buildings [12]. This adherence to object boundaries is superior to arbitrary grid partitions and helps in preserving important spatial structures, which is vital for accurate feature extraction and classification.
- **Generation of Meaningful and Robust Features:** Superpixels inherently group spectrally and spatially similar pixels, thereby capturing local context and reducing within-region variance. This makes them more robust and semantically richer units for feature extraction compared to individual pixels, which often carry limited semantic information in isolation or are more susceptible to noise. Features aggregated over a superpixel (e.g., mean spectral values, textural statistics, or geometric properties as used in this thesis) tend to be more stable and discriminative inputs for machine learning models like GNNs [53]. Superpixels inherently incorporate perceptual information, making them a more practical and semantically rich input for training models tasked with identifying complex structures.
- **Bridging Pixel-Level Detail and Object-Level Context:** The use of superpixels as nodes in a graph allows the GNN to operate on more manageable and semantically coherent units. This approach effectively bridges the gap between the fine-grained spectral information available at the pixel level and the higher-level spatial context that is crucial for identifying built-up objects like buildings. The GNN can then learn complex relationships and contextual dependencies between these superpixel-nodes [57].

In essence, superpixel segmentation transforms the raster image into an intermediate representation that is more aligned with how humans perceive scenes (as a collection of objects or segments) and is more computationally amenable for graph-based deep learning. For the task of building footprint extraction, representing the image as a graph of superpixels allows the GNN to leverage both the intrinsic properties of these segments and their spatial interrelations to make

more informed classification decisions, which is a central tenet of the methodology developed in this thesis.

1.5 Leveraging and Assessing OpenStreetMap Data for Reliable Ground Truth

OpenStreetMap (OSM) stands as a widely recognized, collaboratively maintained, and freely accessible global geospatial database. It has become a cornerstone for various mapping and spatial analysis applications due to its extensive coverage and the richness of its user-contributed data [65]. As a prime example of Volunteered Geographic Information (VGI), OSM relies on a global community of contributors to create, update, and maintain its dataset, which includes a wealth of information about the built environment. Crucially for this research, OSM contains detailed building footprint data for many areas, which is essential for defining the spatial extents of built-up structures and can serve as a valuable source for generating ground truth labels for supervised machine learning models [60].

The building footprint data available within OSM holds significant potential for generating training data and subsequently for assessing and validating the building footprints extracted from Sentinel-2 imagery using the GNN-based approach proposed in this thesis. Given that OSM aims to map real-world features, its building footprint layer can, in principle, serve as a valuable reference. By comparing the spatial extents and locations of building segments identified by the GNN with those present in an OSM dataset, it is possible to derive quantitative measures of the proposed method’s performance.

However, while OSM offers unparalleled accessibility and a vast repository of geospatial information, its nature as a VGI product means that data quality, including completeness, positional accuracy, thematic accuracy, and currency, can vary considerably across different geographic regions and even within the same area [20]. OSM data in certain regions, particularly in less mapped or rapidly changing developing areas, might suffer from incompleteness (not all existing buildings may be mapped) or inaccuracies (mapped building footprints may not perfectly align with their real-world counterparts or may have incorrect attributes) [47]. Therefore, a critical component of this thesis is not just the utilization of OSM data, but also the incorporation of a dedicated methodology for rigorously assessing its quality within the chosen study area of Algiers, Algeria. This assessment is vital to ensure a realistic understanding of OSM’s reliability and to generate trustworthy training labels for the GNN model, as well as to provide a properly contextualized reference for validation.

The core technique employed for this OSM data assessment in this research involves a multi-source cross-validation approach. This compares OSM building footprints against other large-scale, publicly available, and independently generated building footprint datasets, specifically Google Open Buildings (GOB) and the Overture Maps building theme layer [63]. This cross-validation aims to identify areas of consensus (buildings present in multiple datasets, indicating higher reliability) and to flag potential omissions or inaccuracies in OSM (e.g., buildings unique to OSM or missing from it but present in GOB/Overture). Understanding the potential limitations of the OSM data, even after such assessment, is crucial for interpreting the validation results of the GNN-extracted building footprints accurately. Any discrepancies found during the GNN validation process might be attributable to limitations in either the proposed GNN extraction method, the inherent quality of the reference data used for training (derived from assessed OSM), or remaining uncertainties in the validation dataset itself. The detailed methodology for this OSM data assessment and the generation of reliable ground truth labels is a cornerstone of Chapter 3 (specifically Section 3.2.3).

1.6 Research Questions and Objectives

The primary motivation of this Master’s thesis is to explore and evaluate an advanced, data-driven approach for urban building footprint identification, addressing the limitations of traditional methods and leveraging the capabilities of modern remote sensing data and Artificial Intelligence. Specifically, this research focuses on the application of Graph Neural Networks (GNNs) to Sentinel-2 imagery, with a critical emphasis on the role of rigorously assessed OpenStreetMap (OSM) data for generating reliable ground truth within the dynamic urban context of Algiers, Algeria.

To guide this investigation, the following primary research questions have been formulated:

1. How effectively can a Graph Neural Network (GNN) architecture, specifically UrbanGraphSAGE, utilizing a graph structure derived from SLIC superpixel segmentation of Sentinel-2 imagery, extract urban building footprints within the defined study region of Algiers?
2. What is the impact of a systematic OSM data assessment process—Involving multi-source cross-validation with Google Open Buildings and Overture Maps datasets, complemented by temporal stability analysis—on the quality of derived training labels and, consequently, on the performance and reliability of the GNN model for building footprint extraction?

3. To what extent does the application of image-level spectral data augmentation during the training phase enhance the generalization capabilities and overall performance of the UrbanGraphSAGE model?
4. How does the performance of the optimized UrbanGraphSAGE model compare against other established GNN architectures, namely Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), when applied under identical data conditions for urban building footprint identification in Algiers?
5. Which spectral and geometric features derived from Sentinel-2 imagery are most salient for the task of building footprint identification using the UrbanGraphSAGE model in the Algiers urban environment, based on feature importance analysis?

To systematically address these research questions, the following SMART (Specific, Measurable, Achievable, Relevant, Time-bound) objectives were established for this thesis. The detailed methodological steps undertaken to achieve these objectives are presented in Chapter 3.

The primary objectives guiding the research are:

- **To develop and implement a robust data processing pipeline** for acquiring, preprocessing (including band resampling, normalization, and composite creation), and preparing Sentinel-2 imagery and multi-source building footprint vector data (OSM, Google Open Buildings, Overture Maps) for the Algiers study area.
- **To conduct a rigorous quality assessment of OpenStreetMap (OSM) building footprint data** through multi-source cross-validation and temporal stability analysis, aiming to generate a reliable, high-confidence ground truth dataset for GNN model training and evaluation.
- **To design, implement, and train a Graph Neural Network model (UrbanGraphSAGE)** based on superpixel segmentation (using SLIC) of Sentinel-2 imagery, incorporating image-level spectral data augmentation to enhance model robustness.
- **To evaluate the performance of the developed UrbanGraphSAGE model** for building footprint extraction using standard quantitative metrics (F1-Score, Precision, Recall, Accuracy) and qualitative visual assessment, and to compare its performance against alternative GNN architectures (GCN and GAT).

- To analyze the impact of data assessment and spectral augmentation strategies on model performance and to investigate the relative importance of input features for the building footprint identification task.

By addressing these questions and achieving these objectives, this thesis aims to contribute valuable insights and a validated methodological framework for leveraging advanced GeoAI techniques for enhanced urban monitoring.

1.7 Scope and Limitations

This Master's thesis focuses on the development and evaluation of a novel Graph Neural Network (GNN) based methodology for the extraction of urban building footprints from Sentinel-2 satellite imagery, with a significant emphasis on the role of assessed OpenStreetMap (OSM) data in generating reliable ground truth. The research is defined by the following scope:

- **Primary Objective:** The central aim is to extract building footprints, resulting in a binary classification (building/non-building) at the level of superpixels derived from Sentinel-2 imagery.
- **Geographic Focus:** The experimental analysis is conducted within a specifically defined Area of Interest (AOI) in Algiers, Algeria. This region is characterized by significant urban density, ongoing development, and diverse urban morphology, including both formal and potentially informal settlements, providing a challenging and realistic testbed.
- **Satellite Data:** The primary Earth Observation data utilized is Sentinel-2 Level-2A multispectral imagery, valued for its spatial resolution (10-20m effective resolution after resampling for key bands), temporal frequency, and free availability. A 15-channel composite image, including 12 spectral bands and 3 derived indices (NDVI, NDBI, NDSI), serves as the input for superpixel segmentation.
- **Ground Truth Data and Assessment:** OpenStreetMap (OSM) building data serves as the initial reference for ground truth. Its quality is rigorously assessed through a multi-source cross-validation against Google Open Buildings (GOB) and Overture Maps building datasets, further refined by temporal NDVI stability checks. The resulting confidence-scored and temporally verified building footprints are used to generate training and validation labels for the GNN model.

- **AI Model:** The study specifically investigates the application and performance of the UrbanGraphSAGE architecture, a type of GNN, operating on graphs constructed from superpixels generated using the Simple Linear Iterative Clustering (SLIC) algorithm. Comparisons are made with GCN and GAT architectures under similar conditions.
- **Extraction Target:** The definition of "building" is primarily guided by the features explicitly identified as building footprints within the assessed OSM, GOB, and Overture datasets. Extensive sealed surfaces not directly associated with these defined building polygons are not the primary target for extraction.

While this research aims for a rigorous and comprehensive evaluation, certain limitations are acknowledged:

- **Reference Data Variability and Residual Uncertainty:** The completeness, accuracy, and temporal currency of the external building footprint datasets (OSM, GOB, Overture) can vary spatially, even within the study area. While the assessment process aims to mitigate this, residual uncertainties may persist and influence the perceived accuracy and generalization of the GNN model.
- **Definition of Building Footprints:** The reliance on existing building footprint datasets for ground truth means that other components of the urban fabric or very small, unmapped structures might not be explicitly targeted or captured.
- **Superpixel Segmentation Dependency:** The performance of the GNN is inherently linked to the quality of the initial superpixel segmentation. While SLIC is a widely used and optimized algorithm, its parameters were set based on systematic experimentation for this study; an exhaustive optimization across all possible superpixel algorithms and their full parameter spaces was beyond the current scope. Imperfections in superpixel boundaries can affect node feature purity.
- **Temporal Analysis Simplification:** The temporal consistency check for OSM data quality assessment was based on a limited number of Sentinel-2 image dates due to project timeline constraints, rather than a comprehensive, dense time-series analysis.
- **Computational Resources and Model Exploration:** Training and experimenting with GNNs can be computationally intensive. The scope of

hyperparameter tuning for all GNN architectures explored (UrbanGraphSAGE, GCN, GAT) was necessarily constrained by available resources. Exploration of a wider range of GNN architectures or more extensive tuning could yield different results.

- **Model Transferability:** The GNN model trained and optimized for the specific characteristics of the Algiers study area may require further adaptation, fine-tuning, or re-training for optimal performance if applied to different geographical regions with distinct urban morphologies, building styles, or prevalent atmospheric conditions.
- **Context of Feature Importance Analysis:** As noted, the detailed permutation feature importance study was conducted on a non-augmented iteration of the UrbanGraphSAGE model. The relative importance of features might vary for the final model trained with spectral augmentation.

Acknowledging these scope boundaries and limitations is crucial for a balanced interpretation of the research findings and for identifying productive avenues for future work.

1.8 Thesis Structure

The remainder of this thesis is structured as follows:

- **Chapter 2: Literature Review and Theoretical Framework:** This chapter provides a comprehensive overview of the existing body of knowledge relevant to this research. It delves into remote sensing principles for building identification, Sentinel-2 data characteristics, the evolution of AI techniques from traditional machine learning to Deep Learning (CNNs and GNNs), the theoretical foundations of superpixel segmentation (specifically SLIC), and the characteristics and quality assessment considerations for OpenStreetMap data when used as ground truth.
- **Chapter 3: Methodology:** This chapter presents a detailed exposition of the GNN-based methodology developed for extracting building footprints. It covers the study area description, data acquisition (Sentinel-2, OSM, GOB, Overture), extensive preprocessing steps, the OSM data assessment and ground truth generation pipeline, SLIC superpixel segmentation, graph construction with feature engineering, the architecture of the UrbanGraphSAGE model (and comparative GCN/GAT models), training procedures including spectral augmentation, and evaluation protocols.

- **Chapter 4: Results and Discussion:** This chapter presents and analyzes the outcomes of the experimental evaluation. It includes quantitative performance metrics for all trained models, qualitative visual assessments of the extracted building footprints, an analysis of the impact of data assessment and augmentation, findings from SLIC parameterization and feature importance studies, and a discussion contextualizing these results and their implications.
- **Chapter 5: Conclusion:** The final chapter summarizes the key findings and contributions of the thesis. It reiterates the research objectives, discusses the significance of the work, acknowledges its limitations, and offers concluding remarks on the broader implications and potential of the developed framework for urban monitoring.
- **Recommendations for Future Work:** Building upon the discussion in Chapter 4, specific avenues for future research, including explorations of different GNN architectures, integration of additional data sources, and advancements in data augmentation and ground truth generation, are outlined (primarily detailed in Section 4.3.5 and summarized in the overall conclusion).
- **References:** A comprehensive list of all cited literature and data sources.
- **Appendices (Optional):** May include supplementary materials, such as detailed hyperparameter settings not covered in the main text or additional visual results.

Chapter 2

Literature Review and Theoretical Framework

2.1 Introduction

The accurate and timely extraction of building footprints from satellite imagery is a task of increasing importance, driven by diverse applications ranging from urban planning and disaster management to environmental monitoring, as detailed in Chapter 1. While traditional methods for this task face significant limitations in terms of scalability, accuracy, and the ability to handle complex urban environments, advancements in remote sensing technologies, particularly with accessible and high-revisit missions like Sentinel-2, coupled with the evolution of Artificial Intelligence (AI), offer promising new avenues for more efficient and robust solutions.

This thesis proposes and evaluates a novel framework for building footprint extraction that uniquely integrates Graph Neural Networks (GNNs) applied to superpixel-segmented Sentinel-2 imagery. A cornerstone of this framework, and a key focus of this research, is the rigorous assessment and utilization of OpenStreetMap (OSM) data. Specifically, OSM building data is cross-validated with other large-scale datasets, namely Google Open Buildings (GOB) and Overture Maps, to generate reliable ground truth labels for training the GNN model and for its subsequent validation.

This chapter provides the essential theoretical and literary context for these methodological components. It will comprehensively review the existing body of knowledge pertinent to:

1. Remote sensing principles and Sentinel-2 data characteristics relevant to building footprint detection.
2. The application of Artificial Intelligence, from traditional machine learning

to Deep Learning, with a specific focus on the rationale and capabilities of Graph Neural Networks in this domain, contrasting them with Convolutional Neural Networks (CNNs) where appropriate.

3. The role and techniques of superpixel segmentation, particularly Simple Linear Iterative Clustering (SLIC), as an effective preprocessing step for graph-based image analysis.
4. The characteristics of OpenStreetMap data, its inherent quality considerations, robust methods for its assessment (including multi-source cross-validation), and its utilization as ground truth in machine learning applications.
5. The principles and benefits of data augmentation, specifically image-level spectral augmentation, in the context of remote sensing and deep learning.

Through this extensive review, the chapter aims to establish the theoretical underpinnings of the research, identify existing knowledge gaps that this thesis seeks to address, and provide a clear and well-supported justification for the specific methodological choices adopted in the subsequent chapters, particularly Chapter 3.

2.2 Remote Sensing for Building Footprint Extraction

Remote sensing technology provides an invaluable means for observing, monitoring, and analyzing the Earth's surface, with applications spanning numerous disciplines. In the context of urban studies and management, satellite-based remote sensing has become particularly crucial for mapping and characterizing the built environment, including the precise delineation of building footprints [51].

2.2.1 Principles of Satellite-Based Building Footprint Identification

The identification of building footprints from satellite imagery relies on the fundamental principle that different materials on the Earth's surface interact with electromagnetic radiation in distinct ways. Passive satellite sensors, such as the Multi-Spectral Instrument (MSI) aboard the Copernicus Sentinel-2 satellites, measure the solar radiation reflected or emitted from the surface across various spectral bands, ranging from the visible to the short-wave infrared portions of the spectrum [50].

Materials commonly used in building construction—such as concrete, asphalt, metal, glass, and various roofing materials—exhibit unique spectral reflectance characteristics, often referred to as spectral signatures. These signatures are typically different from those of natural land cover types like green vegetation (which strongly reflects in the near-infrared and absorbs in the red), water bodies (which absorb most incident NIR and SWIR radiation), and bare soil [50]. The ability to capture these distinct spectral responses across multiple bands is the primary basis for differentiating buildings from their surroundings using multispectral satellite imagery.

Beyond spectral information, the geometric attributes of buildings also play a significant role in their identification from satellite data. Buildings often exhibit characteristic shapes (e.g., rectangular, L-shaped, polygonal), consistent sizes within certain typologies, and regular orientations, particularly in planned urban areas [28]. The ability to discern these geometric features is, however, contingent upon the spatial resolution of the satellite imagery relative to the scale of the structures being mapped. Advanced image processing techniques, such as the analysis of shadows cast by buildings, can provide additional ancillary cues. Shadows can indicate the presence and approximate height of vertical structures, aiding in their detection and the precise delineation of their footprints, although they can also obscure adjacent features if not properly handled [19].

The methodologies for building footprint extraction from satellite imagery have evolved considerably over time. Early approaches were predominantly manual, involving visual photo-interpretation and digitization of building outlines by human operators. While potentially accurate, these methods are extremely labor-intensive, time-consuming, subjective, and not scalable for large areas or frequent updates [62]. Subsequent advancements led to the development of semi-automated techniques, which often still required significant human intervention for parameter tuning or post-processing. The current research frontier, driven by the increased availability of high-resolution multispectral satellite imagery and progress in machine learning, emphasizes fully automated methods capable of extracting building footprints with greater efficiency, consistency, and accuracy [57]. The development of algorithms that can learn and exploit the complex spectral, spatial, and contextual characteristics indicative of buildings within diverse urban environments is central to this pursuit, moving beyond simple built-up area identification towards the detailed delineation of individual building structures.

2.2.2 Key Characteristics of Sentinel-2 Relevant to Building Footprint Extraction

The Copernicus Sentinel-2 mission, a cornerstone of the European Union's Earth observation program, comprises a constellation of two identical polar-orbiting satellites (Sentinel-2A and Sentinel-2B). These satellites operate in tandem to provide high-resolution, multi-spectral imagery of the Earth's land surfaces, coastal waters, and major islands, making Sentinel-2 a highly valuable and widely utilized data source for a multitude of applications, including urban monitoring and building footprint extraction [22].

Several key characteristics of Sentinel-2 make it particularly relevant and well-suited for the task of building footprint identification, as undertaken in this thesis:

- **Temporal Resolution:** With both satellites operational, the Sentinel-2 mission achieves a high revisit frequency, providing imagery over most land surfaces approximately every 5 days at the equator under cloud-free conditions [22]. This frequent coverage is highly advantageous for monitoring dynamic urban environments like Algiers, where rapid changes can occur. It increases the probability of acquiring cloud-free imagery suitable for analysis and also supports applications requiring time-series data, such as the temporal stability assessment of ground truth features performed in this research (see Section 3.2.3).
- **Spatial Resolution:** The Sentinel-2 Multi-Spectral Instrument (MSI) acquires data at three different native spatial resolutions [22]:
 - Four spectral bands at **10 meters**: Blue (B2, 490 nm), Green (B3, 560 nm), Red (B4, 665 nm), and Near-Infrared (NIR, B8, 842 nm). These 10m bands provide the highest spatial detail and are crucial for resolving individual building structures and urban features.
 - Six spectral bands at **20 meters**: Including four bands in the Vegetation Red Edge spectrum (B5, 705 nm; B6, 740 nm; B7, 783 nm; B8A, 865 nm) and two Short-Wave Infrared bands (SWIR1: B11, 1610 nm; SWIR2: B12, 2190 nm). These bands are critical for vegetation analysis and for distinguishing built-up areas and soil types.
 - Three spectral bands at **60 meters**: Primarily for atmospheric correction and cloud screening (B1 Coastal Aerosol, 443 nm; B9 Water Vapour, 945 nm; B10 SWIR - Cirrus, 1375 nm, though B10 is often not available in L2A products distributed for land applications).

For building footprint extraction at the level of detail pursued in this thesis, the 10-meter bands are of paramount importance. The 20-meter bands, particularly the SWIR bands, also provide critical discriminative information and were resampled to 10 meters in this study to create a consistent multi-channel composite image (as detailed in Section 3.2.2).

- **Spectral Resolution and Capabilities:** The MSI sensor captures data in 13 spectral bands (though typically 12 are used for land applications from L2A products, as B10 is often omitted or used pre-L2A) [22]. This rich spectral information, spanning the visible, near-infrared, red-edge, and short-wave infrared portions of the electromagnetic spectrum, is essential for differentiating various building materials from other land cover types (e.g., vegetation, water, bare soil). For instance, SWIR bands (B11, B12) are particularly effective for distinguishing artificial built-up surfaces from spectrally similar features like dry soil or certain types of rock, due to differences in moisture content and material composition [57]. The Red Edge bands are sensitive to vegetation health and stress, which aids in separating buildings from surrounding green spaces. The combination of these bands allows for the calculation of various spectral indices that further enhance the ability to characterize and identify potential building areas.
- **Data Accessibility and Cost-Effectiveness:** Sentinel-2 data products are provided under a free, full, and open data policy established by the Copernicus program [22]. Data is readily accessible through platforms like the Copernicus Data Space Ecosystem. This open accessibility and cost-effectiveness make Sentinel-2 an exceptionally well-suited data source for research, operational applications, and large-area mapping, removing barriers associated with the cost of commercial high-resolution imagery.

In synergy, Sentinel-2's spectral richness, appropriate spatial resolution for many urban structures (especially when superpixel aggregation is employed), and frequent revisit time provide a robust and regularly updated dataset ideal for building footprint extraction and urban monitoring. The 10-meter resolution of key bands offers the necessary spatial granularity for delineating building outlines when aggregated at the superpixel level, while the broader spectral range, including the SWIR and Red Edge bands, furnishes critical information about material composition and context, facilitating the differentiation of buildings from other land cover types and informing the feature set used by the GNN.

2.2.3 Commonly Used Spectral Indices and Their Relevance to Building Identification

Spectral indices, which are mathematical combinations or transformations of reflectance values from two or more spectral bands, are widely used in remote sensing to enhance specific features or properties of the Earth's surface that may not be readily apparent in individual bands alone. In the context of urban area analysis and preliminary building identification, several indices are frequently employed. For this research, the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), and the Normalized Difference Soil Index (NDSI) were calculated and utilized as input features for the GNN model, alongside the raw spectral bands.

- **Normalized Difference Vegetation Index (NDVI):** NDVI is one of the most established and widely used indices for assessing the presence, density, and vigor of live green vegetation [2]. It is calculated using reflectance values from the Near-Infrared (NIR) and Red portions of the spectrum:

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

Healthy vegetation strongly reflects NIR light and absorbs red light (for photosynthesis), leading to high positive NDVI values (typically ranging from +0.2 to +0.9). Conversely, non-vegetated surfaces such as bare soil, water, snow, and artificial structures (including buildings) typically exhibit low positive, near-zero, or even negative NDVI values [3]. While not a direct indicator of buildings, very low NDVI values within urbanized areas can help differentiate non-vegetated artificial surfaces from surrounding vegetation, thus playing an indirect role in highlighting potential built-up areas.

- **Normalized Difference Built-up Index (NDBI):** NDBI is more specifically designed to highlight built-up areas by exploiting the unique spectral characteristics of artificial surfaces, particularly their higher reflectance in the Short-Wave Infrared (SWIR) region compared to the NIR region [5]. It is typically calculated as:

$$\text{NDBI} = \frac{(\text{SWIR} - \text{NIR})}{(\text{SWIR} + \text{NIR})}$$

For Sentinel-2 imagery, the SWIR band used is typically B11 (SWIR1) or B12 (SWIR2) (both at native 20m resolution, resampled to 10m in this study), and the NIR band is B8 (10m resolution). Built-up areas generally

exhibit positive NDBI values, while vegetation shows low NDBI values, and water bodies have negative values [5]. NDBI is therefore a key index for directly identifying potential concentrations of man-made structures.

- **Normalized Difference Soil Index (NDSI):** To help distinguish bare soil from built-up areas and vegetation, the Normalized Difference Soil Index was included. For this research, it was calculated using the Short-Wave Infrared (SWIR1) and Green bands from the Sentinel-2 imagery. The formula is:

$$NDSI = \frac{(SWIR1 - Green)}{(SWIR1 + Green)}$$

This formulation leverages the characteristic that bare soil tends to have higher reflectance in the SWIR region compared to the Green region. This index is particularly useful for reducing spectral confusion between certain bright soil types and non-vegetated artificial surfaces like building rooftops.

Utility and Limitations of Indices for Building Identification: Both NDVI and NDBI provide valuable initial indicators for locating potential building clusters and serve as discriminative features for machine learning models. Low NDVI and high NDBI values often correspond to built-up areas. However, relying solely on these indices for precise building footprint extraction in complex urban environments presents several limitations [56]:

- **Spectral Confusion:** Built-up areas can share spectral index values similar to other non-building impervious surfaces (e.g., roads, parking lots) and even bare soil or rock outcrops, particularly in arid or semi-arid regions. This can lead to misclassification if NDBI is used as the sole determinant [5].
- **Intra-Urban Variability:** The presence of urban vegetation (e.g., trees overhanging buildings, gardens) can lower NDBI values of actual built-up structures or increase NDVI, complicating their distinction from purely vegetated areas.
- **Material and Shadow Effects:** Variations in building materials (e.g., different roof types and ages), and the presence of shadows, significantly alter the spectral signatures and consequently the index values of buildings, making it difficult to define universal thresholds for these indices that consistently and accurately delineate all building footprints [19].

Therefore, while these spectral indices are useful as part of a richer feature set for characterizing superpixels (as was done in this thesis), their inherent limitations necessitate the use of more advanced techniques, such as the proposed GNN

approach that can learn from context and a wider array of features, to achieve accurate and detailed building footprint extraction.

2.2.4 Essential Preprocessing of Sentinel-2 Data for Urban Building Footprint Extraction

Effective utilization of Sentinel-2 imagery for building footprint extraction, particularly within heterogeneous urban environments, necessitates several essential preprocessing steps beyond the standard Level-2A product generation. These steps ensure data quality, consistency, and reliability for subsequent analysis and model training [51]. While the Copernicus program provides Sentinel-2 data at various processing levels, and this research primarily utilized Level-2A (L2A) products (already providing Bottom-Of-Atmosphere (BOA) surface reflectance and orthorectification [22]), several additional user-end preprocessing considerations remain crucial, as implemented in this thesis (detailed in Section 3.2.2):

1. **Cloud and Cloud Shadow Masking:** This is a fundamental step for any optical satellite imagery analysis. Even with L2A products, residual clouds or, more commonly, their shadows can obscure features or introduce erroneous spectral values. Sentinel-2 products include quality assessment (QA) bands (e.g., the QA60 band and the Scene Classification Layer - SCL) that provide per-pixel information on cloud, cirrus, and cloud shadow probability [38]. These bands are typically utilized to create a mask to identify and exclude pixels contaminated by clouds or affected by their shadows, preventing them from negatively impacting subsequent spectral signature analysis and model training. Accurate masking is vital in urban areas where shadows from tall buildings can also be prevalent.
2. **Resampling of Bands to a Common Resolution:** Sentinel-2 bands are provided at different native spatial resolutions (10m, 20m, and 60m). When features for analysis or GNN model input are derived from a combination of bands with differing resolutions (e.g., using 10m visible/NIR bands alongside 20m SWIR bands for calculating NDBI or as direct multi-channel input), it is necessary to resample all bands to a common, typically the finest, resolution being utilized (10m in this study for building-level detail). Common resampling techniques include nearest neighbor, bilinear interpolation, or cubic convolution, with the choice depending on the nature of the data and the requirements of the subsequent analysis (bilinear interpolation was used in this thesis for continuity) [41]. This step ensures a consistent spatial resolution across all input feature layers for superpixel segmentation and GNN

processing.

3. **Mosaicking and Clipping to Area of Interest (AOI):** If the designated study area spans multiple Sentinel-2 tiles (granules), these tiles must first be mosaicked to create a seamless image covering the entire AOI. Subsequently, the imagery (including all original bands and any derived indices) is precisely clipped to the defined boundaries of the study area. This ensures that the analysis is confined to the region of interest, reduces computational load, and standardizes the geographic extent for all processed data.

These preprocessing steps, alongside normalization and NaN handling (discussed in the context of the methodology in Chapter 3), are prerequisites for achieving reliable building footprint extraction. They ensure the quality, consistency, and comparability of the Sentinel-2 data used for superpixel generation, feature extraction, and ultimately, for training and evaluating the GNN model.

2.3 Artificial Intelligence in Remote Sensing for Feature Extraction

The analysis of remotely sensed imagery has undergone a significant paradigm shift with the increasing integration and sophistication of Artificial Intelligence (AI), Machine Learning (ML), and particularly Deep Learning (DL) techniques [40]. This evolution has moved the field from manual interpretation and traditional image processing towards more automated, data-driven, and powerful analytical capabilities.

2.3.1 The Paradigm Shift: AI, ML, and DL in Earth Observation

Artificial Intelligence, in its broadest sense, refers to the capability of machines to perform tasks that typically require human intelligence, such as learning, problem-solving, and decision-making [49]. Machine Learning is a core subfield of AI that focuses on developing algorithms enabling computer systems to learn patterns and make predictions or decisions from data without being explicitly programmed for each specific task [7]. Deep Learning, a further specialization within ML, employs artificial neural networks with multiple layers (hence "deep" architectures) to automatically learn and extract intricate, hierarchical features directly from raw input data, such as satellite images [23]. This ability to learn features autonomously is a key differentiator from traditional ML approaches, which often rely on hand-crafted features.

These advancements have revolutionized remote sensing data analysis, providing powerful tools for tasks crucial to urban studies and environmental monitoring. This includes automated feature extraction for elements like building footprints, accurate land cover classification, efficient object detection, and effective monitoring of environmental changes over time [40]. The capacity of DL techniques, in particular, to learn complex spectral and spatial patterns directly from large volumes of high-dimensional image data has significantly advanced the potential for extracting detailed and reliable geospatial information, including the precise delineation of built-up structures.

2.3.2 Traditional Machine Learning Approaches for Building Footprint Mapping and Their Limitations

Prior to the widespread adoption of deep learning, various traditional machine learning (ML) algorithms were commonly employed for mapping building footprints and broader urban areas from remote sensing imagery. These approaches, often categorized as pixel-based or object-based (operating on image segments), include methods such as:

- Support Vector Machines (SVMs) [11]
- Random Forests (RF) [27]
- K-Nearest Neighbors (KNN) [14]
- Clustering algorithms like k-means for unsupervised tasks [16]

A hallmark of these traditional ML methods is their heavy reliance on **manually engineered features** extracted from the remote sensing data prior to classification. These features typically encompass [18]:

- **Spectral information:** Raw band reflectance values, vegetation indices (e.g., NDVI), build-up indices (e.g., NDBI), and water indices.
- **Spatial/Textural information:** Features derived from image texture (e.g., Gray-Level Co-occurrence Matrix - GLCM, local binary patterns), and morphological characteristics.
- **Geometric/Shape-based features (often used in Object-Based Image Analysis - OBIA):** Attributes of image segments such as area, perimeter, compactness, and form factor.

While these traditional ML approaches achieved considerable success in various remote sensing applications, they exhibit several inherent shortcomings, particularly when applied to the complex task of precise and large-scale building footprint extraction:

- 1) **Dependence on Feature Engineering:** The performance of traditional ML models is critically dependent on the quality and relevance of the hand-crafted features. Designing effective features requires significant domain expertise and manual effort, can be time-consuming, and may necessitate extensive knowledge of remote sensing principles and the specific spectral and spatial characteristics of buildings within diverse urban environments [25].
- 2) **Limited Generalizability:** Features optimized for one geographic region, sensor type, or time period often do not generalize well to other areas or datasets, frequently requiring re-engineering of features and model retraining for new applications [33].
- 3) **Inability to Learn Hierarchical Features:** Traditional ML algorithms typically operate on these pre-extracted, often "flat," features. They generally lack the inherent ability to automatically learn the intricate hierarchical patterns (from low-level edges and textures to mid-level parts and high-level objects) that characterize complex entities like buildings directly from raw pixel data [17].
- 4) **Challenges with High-Dimensional Data:** While multispectral and multitemporal remote sensing data provide rich information, traditional ML methods can struggle with high dimensionality (many spectral bands or derived features). This can potentially lead to the "curse of dimensionality" or the Hughes phenomenon, where an excess of features relative to the number of training samples can degrade classifier performance [1].
- 5) **Difficulty with Spatial Context and Complex Geometries:** Capturing the complex spatial relationships, contextual information (e.g., a building is usually near a road), and varied geometries necessary for precise building footprint delineation (especially for irregularly shaped or closely spaced buildings) can be challenging for pixel-based traditional ML or even segment-based approaches if the feature set is not sufficiently rich or if the classifier cannot effectively model these higher-order relationships [18].
- 6) **Parameter Tuning:** Many traditional ML algorithms require careful and often extensive parameter tuning to achieve optimal performance, which can

be a non-trivial and time-consuming iterative process [30].

These limitations underscore the need for more advanced approaches capable of automatic feature learning and robustly handling the complexities of building footprint extraction from modern remote sensing data. This provides a strong motivation for exploring deep learning techniques, which are designed to overcome many of these challenges.

2.3.3 Deep Learning and CNNs for Semantic Segmentation of Building Footprints

Deep Learning (DL) has emerged as a transformative paradigm in machine learning, particularly for tasks involving complex, high-dimensional data such as remote sensing imagery. DL models, primarily in the form of deep artificial neural networks, overcome many of the limitations associated with traditional machine learning by enabling the **automatic learning of hierarchical feature representations** directly from raw input data, typically through the use of multiple stacked layers of non-linear processing units [23].

Among the diverse deep learning architectures, **Convolutional Neural Networks (CNNs)** have become the dominant and most successful approach for a wide array of image analysis tasks, including the semantic segmentation of building footprints from satellite and aerial imagery [51]. Semantic segmentation aims to assign a class label (e.g., 'building', 'road', 'vegetation') to each pixel in an image. CNNs are specifically designed to process data with a grid-like topology, such as images. They employ:

- **Convolutional Layers:** These layers apply learnable filters (kernels) across the input image to detect spatial hierarchies of features. These filters automatically learn to identify patterns such as edges and textures in the initial layers, progressing to more complex, object-specific parts and eventually entire objects in deeper layers of the network [13].
- **Pooling Layers (Subsampling):** Typically interspersed with convolutional layers, pooling layers (e.g., max pooling) reduce the spatial dimensions of the feature maps, thereby providing a degree of translation invariance (robustness to small shifts in object position) and reducing the number of learnable parameters, which helps control overfitting.
- **Non-linear Activation Functions:** Functions like the Rectified Linear Unit (ReLU) are integral, applied after convolutional operations, enabling the network to learn complex, non-linear relationships within the data [10].

The advent of CNNs has had a profound impact on the field of building footprint extraction. Their ability to automatically learn discriminative features directly from raw image data has led to significant improvements in accuracy, robustness, and scalability compared to traditional ML methods that rely on hand-crafted features. Several CNN architectures have been successfully adapted and developed for the semantic segmentation of buildings, most notably:

- **Fully Convolutional Networks (FCNs):** Pioneering architectures that replaced the fully connected layers in traditional image classification CNNs with convolutional layers, enabling end-to-end training for dense pixel-wise prediction (semantic segmentation) [24].
- **U-Net and its Variants:** The U-Net architecture is particularly popular and effective for biomedical and subsequently remote sensing image segmentation [26]. It features a symmetric encoder-decoder structure. The encoder part progressively downsamples the input image to capture context while learning a hierarchy of features. The decoder part then gradually upsamples these features back to the original image resolution. Crucially, skip connections link features from the encoder path to corresponding layers in the decoder path. This allows the network to fuse high-resolution, fine-grained features from the encoder with the upsampled contextual features from the decoder, enabling precise localization of object boundaries – a critical aspect for accurate building footprint delineation [26].
- **Other Advanced Architectures:** Many other architectures, like DeepLab (employing atrous convolutions for a wider field of view without increasing parameters) [35], SegNet (efficient memory usage during decoding) [34], and various attention-based models, have also been proposed, often building upon these foundational concepts to further enhance segmentation performance.

These CNN-based approaches have demonstrated state-of-the-art performance in identifying building footprints from diverse remote sensing data sources. However, their inherent design for processing grid-like data presents challenges when seeking to leverage graph-based representations of images, such as those derived from superpixel segmentation, which is a key motivation for exploring GNNs in this thesis.

2.4 Superpixel Segmentation and Graph Neural Networks (GNNs)

The limitations of traditional Convolutional Neural Networks (CNNs) in naturally handling non-Euclidean data structures, such as irregularly shaped image segments and their complex interrelations, motivate the exploration of alternative deep learning architectures. This section delves into superpixel segmentation as a method to transform images into graph-like structures, and then introduces Graph Neural Networks (GNNs) as a class of models specifically designed to operate on such graph data, highlighting their synergistic potential for tasks like building footprint extraction.

2.4.1 Superpixel Segmentation: Concepts and Relevance to Graph-Based Image Analysis

Superpixel segmentation is an image preprocessing technique that groups spatially contiguous pixels sharing similar visual characteristics (e.g., color, intensity, texture in the case of multispectral imagery) into larger, perceptually meaningful regions known as superpixels [4], [42]. Instead of processing an image at the highly granular pixel level, superpixel algorithms partition the image into a set of these "atomic regions" or image primitives. This over-segmentation offers several key benefits for subsequent image analysis, particularly when preparing data for graph-based models like GNNs [12]:

- **Computational Efficiency:** By significantly reducing the number of processing units from potentially millions of pixels to typically thousands of superpixels, the computational complexity of downstream algorithms, including graph construction and GNN operations, can be substantially lowered. This leads to faster processing times and reduced memory requirements, which is particularly crucial when dealing with large remote sensing scenes [53].
- **Boundary Adherence and Preservation of Structure:** Well-designed superpixel algorithms aim to generate segments whose boundaries closely align with the natural contours and edges of objects within the image. This provides a more accurate and perceptually relevant initial segmentation compared to arbitrary grid partitions or overly simplistic pixel groupings, aiding in the preservation of object integrity and fine details [6].
- **Generation of Meaningful Regions and Robust Features:** Superpixels inherently group spectrally and spatially similar pixels, thereby capturing

local context and reducing within-region variance. This makes them more robust and semantically richer units for feature extraction compared to individual pixels, which often carry limited semantic information in isolation and are more susceptible to noise. Features aggregated over superpixels (e.g., mean spectral values, textural statistics) tend to be more stable and discriminative inputs for machine learning models.

- **Adaptability to Irregular Shapes:** Unlike fixed-grid approaches, superpixels can conform to the irregular shapes of natural and man-made objects, making them well-suited for representing complex scenes with diverse feature geometries.

Several algorithms exist for superpixel generation, including Simple Linear Iterative Clustering (SLIC) [12], Felzenszwalb-Huttenlocher (FH) [6], Quick Shift (QS), and Compact Watersheds (CW) [21]. Among these, **SLIC** was chosen for this research due to its favorable balance of computational efficiency, its ability to generate relatively compact and uniform superpixels of a controllable number, and its good adherence to object boundaries, especially with multispectral data. SLIC performs a local k-means clustering in a combined feature space, typically CIELAB color space and XY image coordinates for standard images. For the multispectral Sentinel-2 data used in this thesis, SLIC was adapted to operate on the 15-channel composite image, considering similarity across all spectral dimensions as well as spatial proximity. It initializes cluster centers on a near-regular grid and iteratively refines pixel assignments based on a distance metric that considers both spectral similarity (across all channels) and spatial proximity, resulting in spectrally homogeneous and spatially compact segments [12]. The efficiency and effectiveness of SLIC make it a suitable choice for generating the initial superpixel-based graph representation of the Sentinel-2 imagery, forming the nodes for the GNN.

2.4.2 Graph Neural Networks (GNNs): Theoretical Foundations and Suitability for Graph-Structured Data

Graph Neural Networks (GNNs) represent a specialized and rapidly evolving class of deep neural network architectures explicitly designed to perform inference on data structured as graphs [45]. A graph $G = (V, E)$ is a fundamental data structure comprising a set of nodes (or vertices) V that represent entities, and a set of edges E that denote the relationships or connections between these entities. In the context of this thesis, where images are segmented into superpixels (as discussed in Section 2.4.1), each superpixel is treated as a node in the graph. Edges

are typically defined by the spatial adjacency between these superpixels, thereby encoding the neighborhood structure of the image segments. Both nodes (superpixels) and potentially edges can possess features; for instance, node features can include mean spectral values or textural properties derived from the pixels within each superpixel, as was done in this research [53].

Among the diverse family of GNN architectures, **Graph Convolutional Networks (GCNs)** [29] and **GraphSAGE (Graph Sample and Aggregate) hamilton2017inductive** are prominent variants particularly relevant to this research. The core operation in many GNNs, including GCNs and GraphSAGE, is a form of graph convolution or neighborhood aggregation, which enables a node to update its feature representation (embedding) by aggregating information from its local neighborhood. This is typically achieved through a message passing mechanism [36]:

1. **Message Generation:** Each node in the neighborhood sends a "message" (often a transformation of its current feature vector) to the central node.
2. **Aggregation:** The central node aggregates the incoming messages from its neighbors (e.g., through summation, averaging, or a max operation).
3. **Update:** The aggregated message is combined with the central node's own current feature vector and passed through a neural network layer (often a simple linear transformation followed by a non-linear activation function) to compute its updated feature representation for the next GNN layer.

This iterative process of message passing and feature aggregation, performed across multiple GNN layers, allows information to propagate throughout the graph. Consequently, the learned feature representation for each node can capture not only its intrinsic properties but also complex contextual information derived from its multi-hop neighborhood, effectively encoding both local patterns and broader structural relationships within the graph.

GNNs are inherently well-suited for processing graph-structured data due to several key characteristics [45]:

- **Permutation Invariance/Equivariance:** Their operations are designed to be independent of the arbitrary ordering of nodes in the graph's representation, respecting the inherent nature of graph data.
- **Handling Irregularity:** They can naturally handle the irregular connectivity and varying number of neighbors for each node, a common characteristic of real-world graphs, including superpixel adjacency graphs. This contrasts sharply with architectures like CNNs, which are optimized for regular, grid-like data structures (as discussed in Section 2.3.3).

- **Explicit Relational Modeling:** GNNs explicitly leverage the graph’s topology (the edge structure) in their computations, allowing them to directly model and learn from the relationships between entities. This is crucial for tasks where context and inter-entity dependencies are important.
- **Inductive Capabilities (especially in GraphSAGE):** Some GNN architectures like GraphSAGE are designed to be inductive, meaning they learn functions that can generalize to unseen nodes or even entirely new graphs that were not part of the training set, by learning how to aggregate neighborhood information rather than learning embeddings for specific nodes [hamilton2017inductive](#).

This ability to learn from both node features and the underlying graph structure makes GNNs a powerful tool for analyzing data where inter-entity relationships are as crucial as the attributes of the entities themselves. This is precisely the scenario when representing an image as a graph of interconnected superpixels for tasks like building footprint extraction, where the context and spatial arrangement of image segments are vital for accurate classification.

2.4.3 Synergistic Application of Superpixels and GNNs for Building Footprint Extraction

The combination of superpixel segmentation and Graph Neural Networks offers a powerful and synergistic approach for image analysis tasks, including building footprint extraction from remote sensing data, as pursued in this thesis. This methodology involves transforming the input image (e.g., Sentinel-2 data) into a superpixel graph, where each superpixel forms a node, and edges represent spatial adjacencies or other defined relationships between these superpixels [57].

The key aspects of this synergistic framework are:

1. **Superpixels as Meaningful Graph Nodes:** Superpixels, generated via algorithms like SLIC, encapsulate perceptually homogeneous regions. Treating these as nodes provides a more abstract and semantically richer representation than individual pixels.
2. **Rich Node Features:** Each superpixel node is attributed with a comprehensive feature vector. As implemented in this research, these features include aggregated spectral information (e.g., mean reflectance values across multiple Sentinel-2 bands and derived indices like NDVI, NDBI) and geometric attributes (e.g., area, compactness, bounding box characteristics of the superpixel itself). This provides the GNN with detailed information about the intrinsic properties of each image segment.

3. **Edges Encoding Spatial Context:** Edges connect adjacent superpixel nodes, explicitly encoding the local neighborhood structure of the image. This allows the GNN to learn and leverage spatial context during its message passing and aggregation phases.
4. **GNN for Contextual Learning and Classification:** The GNN (e.g., UrbanGraphSAGE) then operates on this attributed superpixel graph. It learns to classify each superpixel node (e.g., as 'building' or 'non-building') by iteratively aggregating feature information from neighboring superpixels and transforming its own features. This process allows the GNN to consider not just the spectral/geometric properties of an individual superpixel but also the characteristics of its surrounding context.

This superpixel-GNN synergy addresses several key challenges:

- **Computational Scalability:** Processing a graph of a few thousand superpixels is significantly more tractable than processing a graph of millions of pixels, making complex GNN models feasible for large satellite scenes [53].
- **Explicit Modeling of Spatial Context:** The graph structure allows the GNN to explicitly model and learn from the spatial relationships between image segments, crucial for disambiguating spectrally similar but contextually different features (e.g., a bright roof vs. a bright road).
- **Adherence to Object Boundaries:** Superpixels tend to respect natural object boundaries better than pixel grids, potentially leading to more accurate delineation of features like building footprints when the GNN classifies superpixel nodes.
- **Robust Feature Representation:** Node features aggregated over superpixels are generally more robust to noise than individual pixel values.
- **Addressing CNN Limitations for Irregular Data:** This approach directly overcomes the limitations of traditional CNNs in handling non-grid structured data, enabling the explicit modeling of relationships between irregularly shaped image segments.

Existing literature in both computer vision and remote sensing has demonstrated the efficacy of this synergistic approach. Studies have shown that GNNs operating on superpixel graphs can learn powerful and discriminative representations by effectively considering both the local characteristics within each superpixel and the broader contextual relationships between neighboring superpixels, often leading to improved performance in tasks like image classification and semantic segmentation

compared to methods that process pixels independently or only consider limited local context [54]. The application of this framework to Sentinel-2 imagery for building footprint identification, as detailed in this thesis, aims to leverage these strengths for accurate and efficient urban mapping.

2.5 OpenStreetMap (OSM) Data for Ground Truth in Earth Observation

The availability of reliable ground truth data is a cornerstone for training and validating supervised machine learning models in remote sensing applications, including the building footprint extraction task addressed in this thesis. OpenStreetMap (OSM), a global, collaborative, and freely editable mapping project, has emerged as a significant source of Volunteered Geographic Information (VGI) and is increasingly utilized as a potential source of ground truth for various Earth observation tasks [65].

2.5.1 Characteristics of OSM Data: VGI, Building Tags, and Open Licensing

OpenStreetMap operates as a wiki-style map of the world, built and maintained by a vast international community of volunteer contributors, ranging from hobbyist mappers to humanitarian organizations and commercial entities [8]. This collective, crowdsourced effort has resulted in an extensive and continuously expanding database of geospatial information, encompassing a wide array of features such as transportation networks, land use polygons, points of interest, and, crucially for this research, building footprints [20].

The fundamental data model of OSM comprises three core elements:

- **Nodes:** Represent specific point locations on the Earth's surface, defined by latitude and longitude coordinates. They can represent point features (e.g., a tree, a traffic light) or serve as constituent parts of ways and relations.
- **Ways:** Are ordered lists of nodes that form linear features (e.g., roads, rivers) or closed polygons representing areal features (e.g., buildings, lakes, land use areas). For building footprints, closed ways are the primary representation.
- **Relations:** Are used to model more complex geographic features or logical relationships by grouping together multiple nodes, ways, and/or other rela-

tions. Examples include multipolygons (e.g., for buildings with courtyards or multiple disconnected parts) or administrative boundaries.

A key characteristic enabling OSM’s versatility is its flexible **tagging system**. Each node, way, and relation can be associated with an unlimited number of tags, which are key-value pairs (e.g., `key=value`) that describe the attributes of the geographic feature [9]. For the purpose of building footprint extraction, OSM contains a wealth of information encoded within building-related tags. The primary tag for identifying buildings is `building=*`, where the value can specify the type of building (e.g., `building=yes` for a generic building, `building=house` for a residential dwelling, `building=commercial`, `building=industrial`). Additional tags can provide further details, such as the building’s height (`height=*`), number of levels (`building:levels=*`), construction material, or even its name (`name=*`).

A significant advantage of OSM for research and practical applications, including its use in this thesis, is its **open licensing**. The OSM data is licensed under the Open Data Commons Open Database License (ODbL) [15]. This license permits users to freely copy, distribute, transmit, and adapt the data, provided they attribute OpenStreetMap and its contributors and share any derivative works under a compatible license. This open and accessible nature makes OSM an invaluable resource for generating ground truth data for training and validating supervised machine learning models in remote sensing, including the GNN-based building footprint extraction model developed in this study.

2.5.2 Known Data Quality Considerations for OSM Building Footprints

While OpenStreetMap provides an invaluable and openly accessible global geospatial dataset, its nature as a Volunteered Geographic Information (VGI) source means it is subject to several known data quality considerations that must be acknowledged and addressed, particularly when using it as a reference for tasks like building footprint extraction [39]. These considerations, which formed the motivation for the data assessment strategy in this thesis (detailed in Section 3.2.3), include:

- **Completeness:** The spatial coverage and density of mapped features, including building footprints, can vary significantly across different geographic regions and even within them. Urbanized and densely populated areas in developed countries often exhibit higher levels of mapping completeness compared to rural, remote, or less developed regions where the volunteer contributor base may be smaller or less active [47]. This heterogeneity means

that in some areas, a substantial number of existing buildings may not be present in the OSM database (omission errors).

- **Positional Accuracy:** The geographic coordinates of features in OSM are contributed by individuals using diverse methods, ranging from high-precision GPS measurements to on-screen manual digitization from various satellite or aerial imagery basemaps of differing quality and geometric accuracy. This can lead to variations in the planimetric (horizontal) accuracy of the data, with some building footprints being more accurately geolocated and shaped than others [20]. Misalignments with the primary analytical imagery (e.g., Sentinel-2) can occur if these are not carefully addressed during preprocessing or ground truth generation.
- **Thematic Accuracy:** This refers to the correctness and consistency of the tags and attributes assigned to features. For building footprints, while the `building=*` tag is standard, the specific value (e.g., `building=house`, `building=commercial`, `building=yes`) or associated tags (e.g., `height=*`) might be incorrect, inconsistently applied, or missing [20]. Due to the open and collaborative nature of OSM, there is no universal, strictly enforced standardization for all tagging practices, leading to potential variations in interpretation and application by different contributors.
- **Heterogeneity and Consistency:** Data quality and mapping styles can be heterogeneous not only between different regions but also within the same area. This arises from the diverse backgrounds, expertise levels, and regional mapping preferences of the myriad contributors [39]. For example, one part of a city might have highly detailed and accurate building footprints, while another part mapped by different individuals might be less so, or use slightly different tagging conventions.
- **Temporal Accuracy (Currency):** The up-to-dateness of OSM data, including building footprints, depends heavily on the activity level of the local mapping community and the rate of real-world change. Some areas may be frequently updated to reflect new constructions or demolitions, while others might not have been revised for significant periods [32]. This can lead to discrepancies when comparing OSM data with recently acquired satellite imagery, where new buildings might be missing in OSM, or demolished buildings might still be present.

Recognizing these potential data quality issues is crucial when considering OSM data as a source of ground truth for supervised learning in remote sensing applications, such as the GNN-based building footprint extraction central to this thesis.

A thorough assessment of the OSM data quality within the specific study area is therefore a necessary prerequisite.

2.5.3 Importance and Methods of OSM Data Quality Assessment for Building Footprints

Given the inherent data quality characteristics of OpenStreetMap (OSM) stemming from its VGI nature (as discussed in Section 2.5.2), a thorough quality assessment of its building footprint data is of paramount importance before its utilization as ground truth for training and validating supervised machine learning models, such as the Graph Neural Network (GNN) developed in this research. The reliability and ultimate performance of any supervised model are directly contingent upon the accuracy, completeness, and consistency of the ground truth data employed during its training phase. The presence of noisy, incomplete, or inaccurate labels in the training dataset can lead a model to learn erroneous patterns, resulting in suboptimal predictive performance and poor generalization capabilities when applied to new, unseen data. Similarly, robust validation of a model's performance requires a reliable, high-quality benchmark dataset to objectively measure its accuracy and identify areas for improvement.

Various methodologies exist for assessing the quality of OSM data, which can be broadly categorized into intrinsic and extrinsic approaches [39]:

- **Intrinsic Quality Assessment:** These methods evaluate OSM data based on its internal characteristics and metadata, without direct comparison to external reference data. Examples include analyzing the edit history of features (e.g., number of versions, number of contributors), the reputation or experience level of contributing mappers, the density of data, topological consistency (e.g., for road networks ensuring connectivity), and the semantic consistency of tags. While intrinsic indicators can offer valuable insights into mapping activity and potential data maturity, they may not fully capture geometric accuracy or completeness relative to the real world [32].
- **Extrinsic Quality Assessment:** This approach, which is the primary focus for data assessment in this thesis, involves comparing OSM data against one or more independent reference datasets presumed to have higher accuracy or reliability for specific features or regions. For building footprints, such reference data could include authoritative datasets from national mapping agencies (if available and current), commercially produced high-resolution building footprint layers, or, as explored in this research and detailed in Chapter 3, other large-scale, publicly accessible building footprint

datasets generated through different means, such as Google Open Buildings (GOB) and the Overture Maps building theme [63].

The comparison in extrinsic validation is typically performed based on several quality elements, including [20]:

- **Completeness:** Assessing the presence or absence of OSM buildings relative to the reference dataset(s) (i.e., are buildings in the reference also in OSM, and vice-versa?), often quantified as rates of true positives, false positives (commission errors), and false negatives (omission errors).
- **Positional Accuracy:** Measuring the geometric agreement between corresponding building footprints in OSM and the reference data, for example, by calculating the offset distance between centroids or using overlap metrics like Intersection over Union (IoU).
- **Shape Fidelity:** Assessing how well the shape of OSM building polygons matches the shape in the reference data.
- **Thematic Accuracy:** Verifying the correctness of tags associated with OSM buildings against reference attributes, if available.

By quantifying these aspects of data quality through extrinsic validation, particularly by cross-comparing OSM with GOB and Overture Maps as performed in this thesis (see Section 3.2.3), a more objective and nuanced understanding of the suitability of OSM data for use as ground truth can be achieved. This allows for informed decisions regarding data filtering, label generation strategies, and the interpretation of the GNN model’s performance.

2.5.4 Utilizing Google Open Buildings and Overture Maps for OSM Cross-Validation

To achieve a robust assessment of the OpenStreetMap (OSM) building footprint data within the Algiers study area, particularly concerning its completeness and positional accuracy, this research employed a cross-validation methodology utilizing other publicly available, large-scale building footprint datasets. Two prominent datasets selected for this purpose, due to their coverage of Algeria, open licensing, and independent generation or aggregation processes, were Google Open Buildings (GOB) and the Overture Maps building theme layer.

Google Open Buildings (GOB) is an extensive open dataset providing building footprints automatically extracted from high-resolution satellite imagery through the application of deep learning techniques by Google AI [46]. Given its

primary coverage focus on regions including Africa, GOB offers a valuable, independently generated source of building footprint information for Algeria, including the study area. Its machine-generated nature provides a different perspective compared to the VGI approach of OSM, making it highly suitable for identifying potential omissions in OSM or areas of disagreement.

Overture Maps Building Data is another significant global open map initiative, aiming to provide a comprehensive and harmonized dataset of various map features, including building footprints, by integrating data from multiple sources [55]. This includes a version of OpenStreetMap data alongside contributions from major commercial entities (like Meta and Microsoft) and community mapping organizations. The Overture Maps data is also openly licensed (e.g., CDLA Permissive v2, ODbL). While Overture incorporates OSM, comparing the specific OSM data snapshot used in this thesis with the Overture building layer (which might use a different OSM vintage or have undergone conflation and quality control) allows for an assessment of how OSM features align with a professionally curated, conflated product. This can highlight areas of OSM that are well-corroborated by other sources within Overture, or areas where Overture provides additional or differing building information.

By systematically comparing the OSM building footprints within the Algiers study area against the corresponding footprints available in the GOB and Overture Maps datasets (as detailed in Section 3.2.3), a more comprehensive understanding of OSM’s quality can be achieved. This cross-validation focuses on:

- Identifying buildings present in OSM but potentially missing in GOB/Overture (potential “phantom” OSM features, unique OSM contributions, or newer features not yet in GOB/Overture).
- Identifying buildings present in GOB/Overture but not in OSM (indicating potential OSM omissions or areas needing updates).
- Assessing positional agreement and geometric similarity where buildings are represented in multiple datasets (e.g., using overlap metrics to quantify consensus).

This process of cross-validation using these complementary large-scale datasets aims to provide a more reliable and nuanced assessment of OSM’s strengths and weaknesses as a source for generating ground truth data to train and validate the GNN model in this thesis, moving beyond reliance on a single, potentially flawed, source.

2.5.5 Critical Role of OSM Data Assessment for GNN Model Reliability

A rigorous assessment of OpenStreetMap (OSM) building footprint data quality, particularly through the multi-source cross-validation with datasets like Google Open Buildings (GOB) and Overture Maps as planned and executed in this thesis, is an absolutely crucial prerequisite for generating reliable training and validation datasets for the supervised Graph Neural Network (GNN) model. The performance and trustworthiness of any supervised machine learning model are fundamentally dependent on the quality of the data it learns from.

1. **Impact on Model Training:** The GNN model learns to map input features derived from Sentinel-2 imagery (via superpixels) to output labels representing building footprints. If the ground truth labels used for this training process are noisy (e.g., incorrect geometries due to poor digitization), incomplete (e.g., missing actual buildings), or inaccurate (e.g., poor positional alignment of footprints relative to the imagery), the GNN will inevitably learn these errors and biases.
 - For instance, if the OSM-derived data used for training is incomplete and misses a significant number of actual buildings in the study area, the trained GNN model will likely underpredict building footprints, failing to identify legitimate structures because it was not shown enough examples of them or learned that such areas are non-building.
 - Similarly, if the positional accuracy of the OSM building footprints is low relative to the Sentinel-2 imagery, the model might learn to associate image features from one location with ground truth labels that are spatially offset. This leads to imprecise and spatially shifted building footprint extractions by the GNN.

Therefore, a comprehensive quality assessment, leading to the selection or refinement of high-confidence building footprints (as was done in this thesis through confidence scoring and temporal stability checks), is necessary to ensure that the training data fed to the GNN model is as accurate and representative of the true building landscape as possible. This directly influences the model's ability to learn discriminative patterns effectively and reduces the risk of learning from erroneous labels.

2. **Ensuring Robust Model Validation:** Beyond training, a reliable and representative validation dataset, also derived from the most trustworthy subset of the assessed building data, is essential for objectively evaluating the

performance of the trained GNN model. Without a high-quality validation set that accurately reflects the real-world distribution and characteristics of building footprints, it is impossible to accurately gauge how well the model is likely to perform on unseen real-world data. Such a benchmark is critical for making informed decisions about model architecture, hyperparameter tuning, identifying model weaknesses, and understanding its true capabilities and limitations. If the validation data itself is flawed, then metrics like F1-score, precision, and recall lose their meaning as true indicators of model performance.

By rigorously assessing the quality of available building footprint data (OSM, GOB, Overture) and using this understanding to inform the generation of training and validation labels, this research aims to:

- Make informed decisions about data cleaning, filtering, and conflation (e.g., prioritizing superpixels labeled based on buildings with high cross-source agreement for training).
- Mitigate the impact of known VGI data limitations (incompleteness, positional inaccuracies) on the GNN’s learning process.
- Provide a more credible and robust evaluation of the GNN’s performance for building footprint extraction, grounded in the best available understanding of the reference data’s quality.

In essence, the reliability and scientific validity of the entire research endeavor, particularly the conclusions drawn about the GNN’s effectiveness, hinge significantly upon the quality of the ground truth data used. Thus, the thorough assessment of building footprint data from OSM, complemented by GOB and Overture, is not merely a preliminary step but an indispensable component of the methodology for developing and evaluating a trustworthy GNN model for urban applications.

2.6 Conclusion

This chapter has provided a comprehensive review of the literature and theoretical concepts underpinning the methodologies employed in this thesis for urban building footprint extraction from Sentinel-2 imagery using Graph Neural Networks (GNNs) and assessed OpenStreetMap (OSM) data. The review has spanned several critical domains, from the fundamentals of remote sensing and satellite data characteristics to the evolution of Artificial Intelligence techniques and the nuances of utilizing volunteered geographic information for scientific research.

The principles of satellite-based building footprint identification were discussed, highlighting how spectral and geometric properties captured by sensors like Sentinel-2’s Multi-Spectral Instrument (MSI) enable the differentiation of built-up structures. The key characteristics of Sentinel-2 data—its temporal, spatial, and spectral resolutions, along with its open data policy—were examined, underscoring its suitability for urban monitoring and as a primary data source for this research. Commonly used spectral indices such as NDVI, NDBI, and NDSI were reviewed for their relevance in characterizing land cover, alongside a discussion of essential preprocessing steps like cloud masking and band resampling necessary to prepare Sentinel-2 data for robust analysis.

The chapter traced the paradigm shift in remote sensing feature extraction, from traditional machine learning approaches reliant on hand-crafted features to the rise of Deep Learning. The capabilities and inherent limitations of Convolutional Neural Networks (CNNs) for semantic segmentation were outlined, particularly their constraints when dealing with non-Euclidean or graph-structured data. This set the stage for introducing superpixel segmentation, with a focus on the SLIC algorithm, as an effective technique for transforming raster imagery into a graph-based representation by creating perceptually meaningful image primitives.

Subsequently, the theoretical foundations of Graph Neural Networks were explored, emphasizing their inherent suitability for learning from graph-structured data by leveraging both node features and topological relationships. Architectures like GCNs and GraphSAGE were introduced, highlighting their message-passing mechanisms and ability to model complex contextual information. The synergistic application of superpixels (as graph nodes) and GNNs was presented as a powerful approach for image analysis, particularly for tasks like building footprint extraction where spatial context is paramount.

Finally, the chapter delved into the characteristics of OpenStreetMap (OSM) data, recognizing its immense value as a VGI source while also critically examining its known data quality considerations, including issues of completeness, positional accuracy, and temporal currency. The importance of rigorous OSM data quality assessment was emphasized, and the methodology of using extrinsic cross-validation with independent, large-scale datasets like Google Open Buildings (GOB) and Overture Maps was introduced as a robust strategy to enhance the reliability of OSM-derived ground truth. The critical role of this assessed ground truth for training trustworthy GNN models was underscored.

Collectively, the literature and concepts reviewed in this chapter establish a strong theoretical and contextual basis for the research undertaken in this thesis. They justify the selection of Sentinel-2 imagery, the superpixel-GNN framework (specifically UrbanGraphSAGE), the meticulous OSM data assessment strategy,

and the use of spectral data augmentation. The identified gaps and challenges in existing approaches further motivate the novel methodological contributions detailed in Chapter 3, aiming to advance the state-of-the-art in automated urban building footprint extraction.

Chapter 3

Methodology

This chapter provides a comprehensive, step-by-step account of the methodology employed for advancing urban building footprint identification within a selected Area of Interest (AOI) in Algiers, Algeria. The approach leverages the synergistic capabilities of Sentinel-2 multispectral imagery, Graph Neural Networks (GNNs), and a rigorously assessed OpenStreetMap (OSM) dataset for ground truth generation. The methodology encompasses several key stages: (1) data acquisition and meticulous preprocessing of both satellite imagery and vector footprint data; (2) a multi-source cross-validation and temporal stability assessment of OSM data to create reliable ground truth labels; (3) superpixel segmentation of the satellite imagery to form meaningful image primitives; (4) construction of a graph representation where superpixels are nodes and spatial adjacencies are edges, with nodes attributed a rich set of spectral and geometric features; (5) the design, training, and optimization of the proposed UrbanGraphSAGE model architecture, including the integration of image-level spectral data augmentation; and (6) the establishment of robust evaluation protocols for assessing model performance and comparing it against alternative GNN architectures. Each stage and the choices made therein are thoroughly justified to ensure scientific rigor, reproducibility, and a clear understanding of the framework's contributions to automated urban feature extraction.

3.1 Problem Formulation

This section reiterates the core research objectives and contextualizes the chosen methodological framework within the broader challenges of accurate and efficient urban monitoring, particularly in dynamic and morphologically diverse environments such as Algiers.

The primary objective of this research is to develop, implement, and critically

evaluate a novel Graph Neural Network (GNN) based framework designed for the precise and efficient identification of urban building footprints from Sentinel-2 satellite imagery. Specifically, the study aims to assess the effectiveness of the UrbanGraphSAGE architecture for the semantic segmentation of urban building footprints. A significant aspect of this investigation is the impact of integrating quality-assessed OpenStreetMap (OSM) data, cross-validated with Google Open Buildings (GOB) and Overture Maps Building Data, as a robust source of ground truth for GNN training and validation. This data-centric aspect is crucial, as the quality of training data fundamentally dictates the performance and reliability of any supervised learning model.

Furthermore, the research evaluates the contribution of image-level spectral data augmentation techniques to enhance the GNN model's robustness and generalization capabilities, particularly in mitigating the effects of variations in atmospheric conditions, illumination, and seasonal changes inherent in satellite imagery. The performance of the proposed UrbanGraphSAGE model, enhanced with these data strategies, is systematically compared against other prominent GNN architectures, specifically Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). These comparative models are trained under similar conditions, including the use of the same assessed ground truth and data augmentation strategies, to ensure a fair and insightful evaluation of architectural merits.

Key research questions guiding this methodology include:

- **GNN Efficacy for Footprint Extraction:** How effectively does the UrbanGraphSAGE GNN architecture, which processes superpixel-derived features and models inter-superpixel neighborhood relationships from Sentinel-2 imagery, perform in the task of identifying urban building footprints within the complex urban landscape of Algiers?
- **Impact of Ground Truth Quality:** To what extent does the utilization of a multi-source, quality-controlled OSM ground truth dataset improve the accuracy, reliability, and generalization of the GNN model for building footprint extraction compared to using raw or unassessed OSM data?
- **Contribution of Spectral Augmentation:** What is the quantitative impact of applying image-level spectral data augmentation during the training phase on the overall performance and robustness of the UrbanGraphSAGE model when evaluated on unseen test data?
- **Comparative GNN Performance:** How does the performance of the optimized UrbanGraphSAGE model compare against alternative GNN ar-

chitectures, such as GCN and GAT, when all models are applied to the same building footprint extraction task using the identical assessed ground truth and data augmentation strategies?

- **Feature Salience in Algiers:** Based on feature importance analysis conducted on the trained UrbanGraphSAGE model, what are the most salient spectral and geometric features derived from Sentinel-2 imagery that contribute most significantly to the task of building footprint identification in the specific and diverse urban context of Algiers?

Accurate and up-to-date urban building footprint maps are crucial for effective urban planning, disaster management, infrastructure development, and monitoring of Sustainable Development Goals (SDGs) [48]. Algiers, as a dynamic and diverse urban environment, presents unique challenges. Its urban morphology ranges from the historic, densely packed Casbah with its organic fabric of narrow alleys and irregular building patterns, to modern, planned areas with wide boulevards and contemporary housing projects. This inherent diversity implies that rigid, traditional feature extraction or pixel-based classification methods might struggle to generalize effectively across such varied contexts. The GNN approach, operating on adaptively shaped superpixels and their relationships, offers a more flexible representation. Superpixels can conform to irregular building shapes, and the GNN can learn contextual features from neighboring superpixels, potentially distinguishing between spectrally similar but morphologically distinct urban fabric patterns. This adaptability is critical for accurately delineating diverse building footprints, particularly in distinguishing informal settlements or older, less geometrically regular structures from non-building areas.

The methodology detailed in this chapter is structured to address these challenges. It begins with the acquisition and rigorous preprocessing of Sentinel-2 imagery and multi-source building footprint datasets (OSM, GOB, Overture Maps) pertinent to the Algiers study area. This is followed by a comprehensive OSM data quality assessment and temporal stability analysis to generate a refined and reliable ground truth mask. Subsequently, superpixel segmentation using the SLIC algorithm is applied to the preprocessed Sentinel-2 composite image. A graph is then constructed, where superpixels form nodes endowed with spectral and geometric features, and edges represent spatial adjacencies. The chapter culminates in the detailed exposition of the UrbanGraphSAGE model architecture, the training procedures (which incorporate image-level spectral data augmentation), and the robust evaluation protocols used to assess its performance and compare it against the GCN and GAT baseline models.

3.2 Data Acquisition and Preprocessing

This section details the acquisition, preparation, and rigorous quality assessment of all geospatial data utilized in this study. The primary objective of this stage was to assemble a consistent and reliable set of input features from Sentinel-2 satellite imagery and to generate a high-quality ground truth mask representing building footprints within the Algiers study area, suitable for training and evaluating the GNN models.

3.2.1 Study Area

The study area for this research is a selected rectangular region within Algiers, the capital city of Algeria. Algiers presents a compelling case study due to its dynamic urban growth and diverse morphological characteristics, ranging from the historic, organically evolved Casbah to modern, planned districts and informal settlements, as discussed in Section 3.1. The specific Area of Interest (AOI) is geographically defined by the following WGS84 (EPSG:4326) coordinates:

```
[[[3.058662, 36.683754],  
 [2.88311, 36.683754],  
 [2.88311, 36.767587],  
 [3.058662, 36.767587],  
 [3.058662, 36.683754]]]
```

The data was reprojected to the Universal Transverse Mercator (UTM) Zone 31N (EPSG:32631), which is the native projection of the Sentinel-2 data used, this bounding box corresponds to approximately (xmin: 489556.67, ymin: 4059792.71, xmax: 505241.05, ymax: 4069096.92) in meters, encompassing an area of roughly 145.8 square kilometers. This AOI was selected to encapsulate a representative mix of urban typologies prevalent in Algiers.

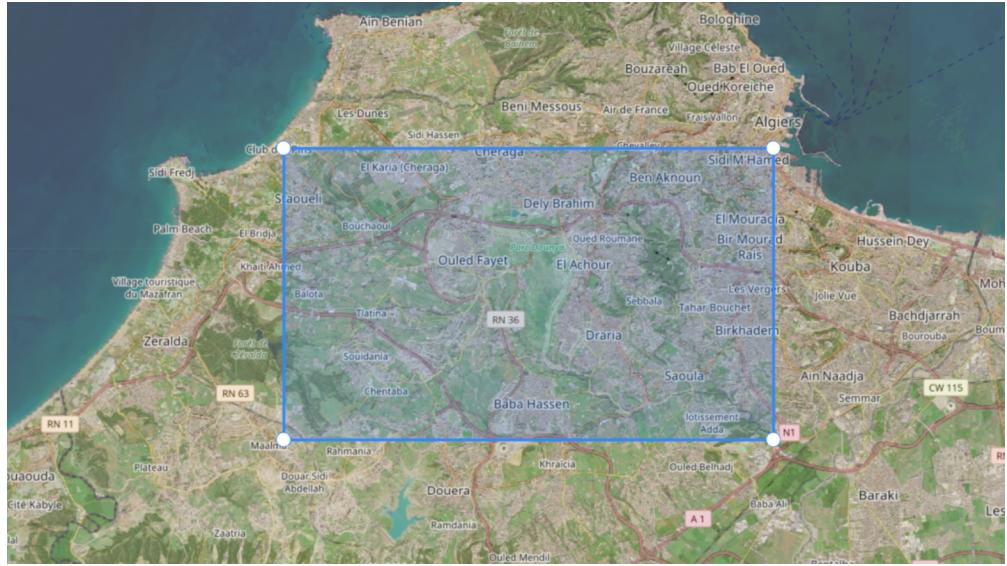


Figure 3.1: Map illustrating the selected Area of Interest (AOI) in Algiers, Algeria, used for this study. The rectangular AOI is shown by the blue overlay. The boundary coordinates defining this AOI are provided in Section 3.2.1. (Base map data source: OpenStreetMap contributors)

3.2.2 Satellite Imagery Acquisition and Preprocessing

Sentinel-2 Data Acquisition

The primary Earth Observation data source for this research was Sentinel-2 multispectral imagery, acquired from the Copernicus Open Access Hub (now Copernicus Data Space Ecosystem). Level-2A products were specifically chosen as they provide Bottom-Of-Atmosphere (BOA) surface reflectance values, having already undergone radiometric, geometric, and atmospheric corrections by the European Space Agency (ESA) [22]. This significantly simplifies the initial preprocessing workflow and ensures that the spectral data are more directly comparable across different acquisitions and suitable for quantitative analysis. Imagery for the Algiers AOI was selected to be as cloud-free as possible, focusing on acquisitions that would support both the primary feature extraction and the temporal stability assessment of ground truth data (details on dates for temporal analysis are in Section 3.2.3).

Sentinel-2 Data Preprocessing Pipeline

A comprehensive preprocessing pipeline was implemented to prepare the Sentinel-2 data for superpixel segmentation and feature extraction. This pipeline, primarily executed using Python libraries such as Rasterio, NumPy, and Scikit-image, involved the following key steps:

- **Band Selection and Composite Image Creation:** A set of 12 Sentinel-2 spectral bands was utilized: B01 (Coastal Aerosol, 60m), B02 (Blue, 10m), B03 (Green, 10m), B04 (Red, 10m), B05 (Vegetation Red Edge 1, 20m), B06 (Vegetation Red Edge 2, 20m), B07 (Vegetation Red Edge 3, 20m), B08 (Near-Infrared - NIR, 10m), B8A (Narrow NIR, 20m), B09 (Water Vapour, 60m), B11 (Short-Wave Infrared 1 - SWIR1, 20m), and B12 (Short-Wave Infrared 2 - SWIR2, 20m). In addition to these reflectance bands, three common spectral indices known for their utility in distinguishing land cover types were calculated:

- Normalized Difference Vegetation Index (NDVI) [2]:

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

- Normalized Difference Built-up Index (NDBI) [5]:

$$\text{NDBI} = \frac{(\text{SWIR1} - \text{NIR})}{(\text{SWIR1} + \text{NIR})}$$

- Normalized Difference Soil Index (NDSI):

$$\text{NDSI} = \frac{(\text{SWIR1} - \text{Green})}{(\text{SWIR1} + \text{Green})}$$

These 12 bands and 3 indices were stacked to form a 15-channel multispectral composite image, which served as the input for subsequent super-pixel segmentation.

- **Cloud and Cloud Shadow Masking:** Although Level-2A products undergo initial cloud screening, residual cloud cover and shadows can persist. To ensure data quality, the QA60 band (providing cloud and cirrus flags) and the Scene Classification Layer (SCL) band (classifying pixels into categories like cloud, shadow, vegetation, water, etc.) were utilized [38]. Pixels identified as contaminated by clouds or their shadows were masked out from further analysis. This step is crucial as clouds and shadows significantly alter spectral signatures and can lead to erroneous feature extraction and classification.
- **Resampling to Uniform Resolution:** Given the varying native spatial resolutions of the selected Sentinel-2 bands (10m, 20m, and 60m), all bands were resampled to a uniform spatial resolution of 10 meters. This was achieved using bilinear interpolation via the `rasterio.warp.reproject`

function, as demonstrated in the SLIC parameter experimentation setup [41]. The 10m resolution was chosen to leverage the highest spatial detail available from Sentinel-2 for delineating building features while maintaining computational tractability for the GNN.

- **Mosaicking and Clipping:** If the AOI spanned multiple Sentinel-2 tiles, the preprocessed tiles would be mosaicked to create a single, seamless image. Subsequently, the entire 15-channel imagery stack was precisely clipped to the defined Algiers AOI boundaries using the reprojected AOI polygon (from UTM Zone 31N) with the `rasterio.mask.mask` function. This ensured that all analyses were confined to the region of interest, optimizing processing load.
- **Normalization:** To ensure that no single band or index disproportionately influenced subsequent processes (like superpixel segmentation or GNN training) due to differing value ranges, all 15 channels of the composite image were normalized. Reflectance bands (B01-B12, B8A), originally in a range potentially up to 10000 for Level-2A surface reflectance, were scaled. This involved dividing by a maximum reflectance value (e.g., 5000, as explored, to handle potential saturation and sensor variations) and then clipping values to the [0, 1] range. Spectral indices (NDVI, NDBI, NDSI), which inherently range from -1 to 1, were normalized to the [0, 1] range using the formula: $\text{Index}_{\text{norm}} = (\text{Index}_{\text{value}} + 1.0)/2.0$. This transformation preserves relative differences while standardizing the input range for the model.
- **NaN Handling:** Any remaining NaN (Not a Number) values within the 15-channel composite image, which could arise from masking operations or edge effects during clipping, were replaced with 0.0 using `numpy.nan_to_num`. This step is critical for the stability of subsequent algorithms like SLIC and GNNs, which typically do not handle NaN inputs gracefully. The resulting 15-channel NumPy array, with dimensions (height, width, 15), served as the final input for superpixel segmentation.



Figure 3.2: Conceptual diagram illustrating the key steps in the Sentinel-2 data preprocessing pipeline, from raw Level-2A product acquisition to the final 15-channel normalized composite image ready for superpixel segmentation. (Source: Author)

3.2.3 Ground Truth Data Acquisition and Assessment Strategy

The generation of a high-quality, reliable ground truth dataset representing urban building footprints is paramount for the successful training and robust evaluation of supervised deep learning models. This study employed a meticulous multi-stage strategy, combining data from OpenStreetMap (OSM) with independent large-scale building datasets (Google Open Buildings and Overture Maps Building Data) for comprehensive cross-validation, followed by a temporal stability assessment to further refine the labels. This approach aimed to mitigate known inconsistencies and incompleteness in individual open datasets and to produce a definitive ground truth mask for the Algiers AOI.

Primary and Reference Building Footprint Datasets

- **OpenStreetMap (OSM) Data:** Building footprint data for the Algiers AOI were acquired from OpenStreetMap using the Overpass API. A custom Python function, `fetch_osm_buildings`, was developed to query for way and relation elements tagged with `building=*`. This function incorporated error handling, retries, and delays to manage API rate limits and ensure successful data retrieval. OSM served as the initial, primary source of building footprint geometries due to its global coverage, open licensing (ODbL), and community-driven update mechanism, offering a readily available, albeit potentially variable quality, dataset.

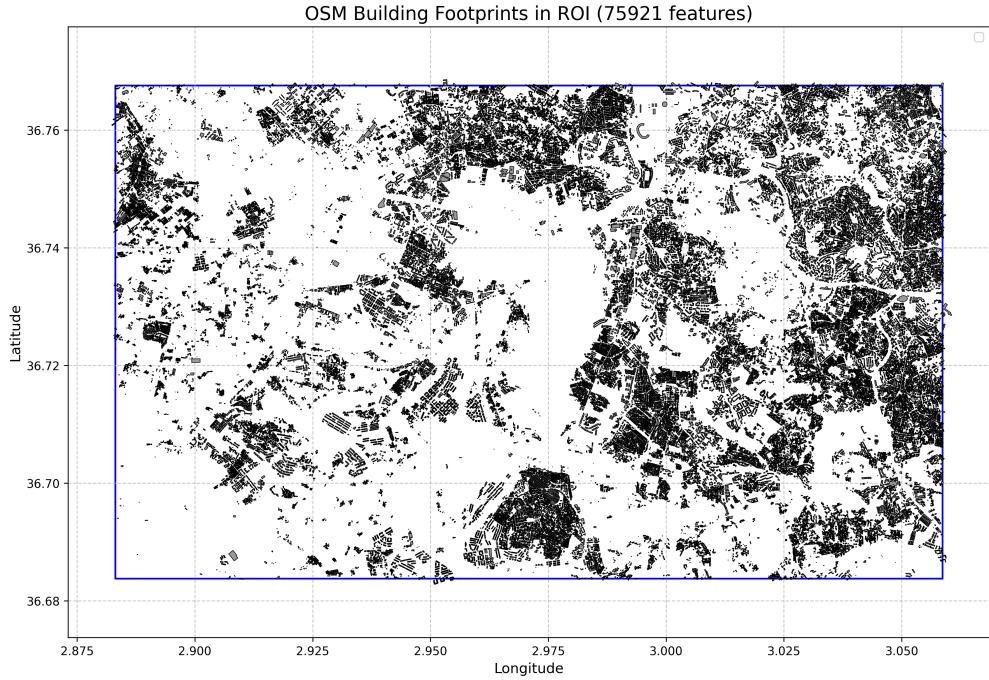


Figure 3.3: OpenStreetMap (OSM) building footprints retrieved for the defined Area of Interest (ROI) in Algiers prior to quality assessment and cross-validation. The ROI boundary is shown in blue. The plot shows 75921 OSM building features within the AOI. (Source: OpenStreetMap contributors, data fetched via Overpass API and plotted by author)

- **Google Open Buildings (GOB):** The Google Open Buildings v3 dataset was utilized as a key independent reference. This dataset provides building footprints automatically extracted from high-resolution satellite imagery using deep learning techniques, with extensive coverage in Africa, including Algeria. Data for the AOI were acquired in GeoParquet format. GOB's independent generation process and large-scale coverage make it an excellent resource for cross-validating and augmenting OSM data.
- **Overture Maps Building Dataset:** The Overture Maps Foundation's building theme layer served as the second independent reference dataset. Overture Maps aims to provide comprehensive and harmonized global map data by aggregating information from multiple sources, including a version of OSM, as well as contributions from commercial and community mapping organizations. Its open licensing and focus on high-quality, conflated data offer another valuable perspective for assessing OSM building footprints.

Table 3.1: Key Characteristics of Ground Truth Data Sources Utilized

Dataset Name	Source/Origin	Key Characteristics	Initial Role in Methodology
OpenStreetMap (OSM)	Crowdsourced (VGI)	Community-contributed, open, global	Primary initial source
Google Open Buildings (GOB)	Google AI	AI-derived from high-res imagery	Cross-validation reference
Overture Maps Buildings	Overture Maps Foundation	Aggregated, multi-source, open	Cross-validation reference

OSM Data Assessment: Multi-Source Cross-Validation

To address potential inconsistencies, incompleteness, or inaccuracies inherent in any single open dataset, particularly OSM, a multi-source cross-validation methodology was implemented. This process aimed to generate a more reliable and robust set of building footprints by leveraging the complementary strengths of OSM, GOB, and Overture Maps.

- **Data Preparation and CRS Alignment:** All three datasets (OSM, GOB, Overture) were clipped to the Algiers AOI and reprojected to a common Coordinate Reference System (CRS), EPSG:4326 (WGS84), to ensure accurate spatial comparison. Invalid geometries were addressed using a `buffer(0)` operation.
- **Spatial Joins:** The core of the cross-validation involved performing spatial joins between pairs of these datasets using GeoPandas (`gpd.sjoin`) with an 'intersects' predicate. This identified building footprints that had any spatial overlap across the different sources. Joins were performed for: OSM-GOB, OSM-Overture, and GOB-Overture.
- **Confidence Scoring:** Based on the results of the spatial joins, a confidence score (ranging from 1 to 3) was assigned to each unique building footprint geometry present in the combined dataset (after initial deduplication by geometry).
 - **Score 3 (High Confidence):** Building footprint present (i.e., spatially intersects) in all three datasets (OSM, GOB, and Overture).
 - **Score 2 (Medium Confidence):** Building footprint present in any two of the three datasets.

- **Score 1 (Low Confidence):** Building footprint present in only one dataset.

This scoring system provided a quantitative measure of inter-dataset agreement for each building.

- **Deduplication:** After assigning confidence scores, a final deduplication based on geometry was performed. If multiple source footprints represented the same physical building (i.e., their geometries overlapped significantly), the footprint with the highest confidence score was retained. This ensured that each distinct physical building was represented only once in the assessed dataset.
- **Identification of Discrepancies:** This process also facilitated the identification of potential "phantom" buildings (those present only in OSM, thus having a confidence score of 1 and a source of OSM) and "missing" buildings (those present in GOB and/or Overture but not in OSM). The analysis showed that a significant number of score 1 buildings originated from the GOB dataset, highlighting its broader coverage compared to OSM and Overture in certain parts of the AOI. Specifically, out of approximately 149,286 unique buildings with a confidence score of 1 from the initial sample, a large proportion (139,739) were solely from GOB.

This multi-source assessment was critical for mitigating biases and errors inherent in any single dataset, leading to a more comprehensive and reliable representation of actual building footprints for subsequent temporal analysis and ground truth generation.

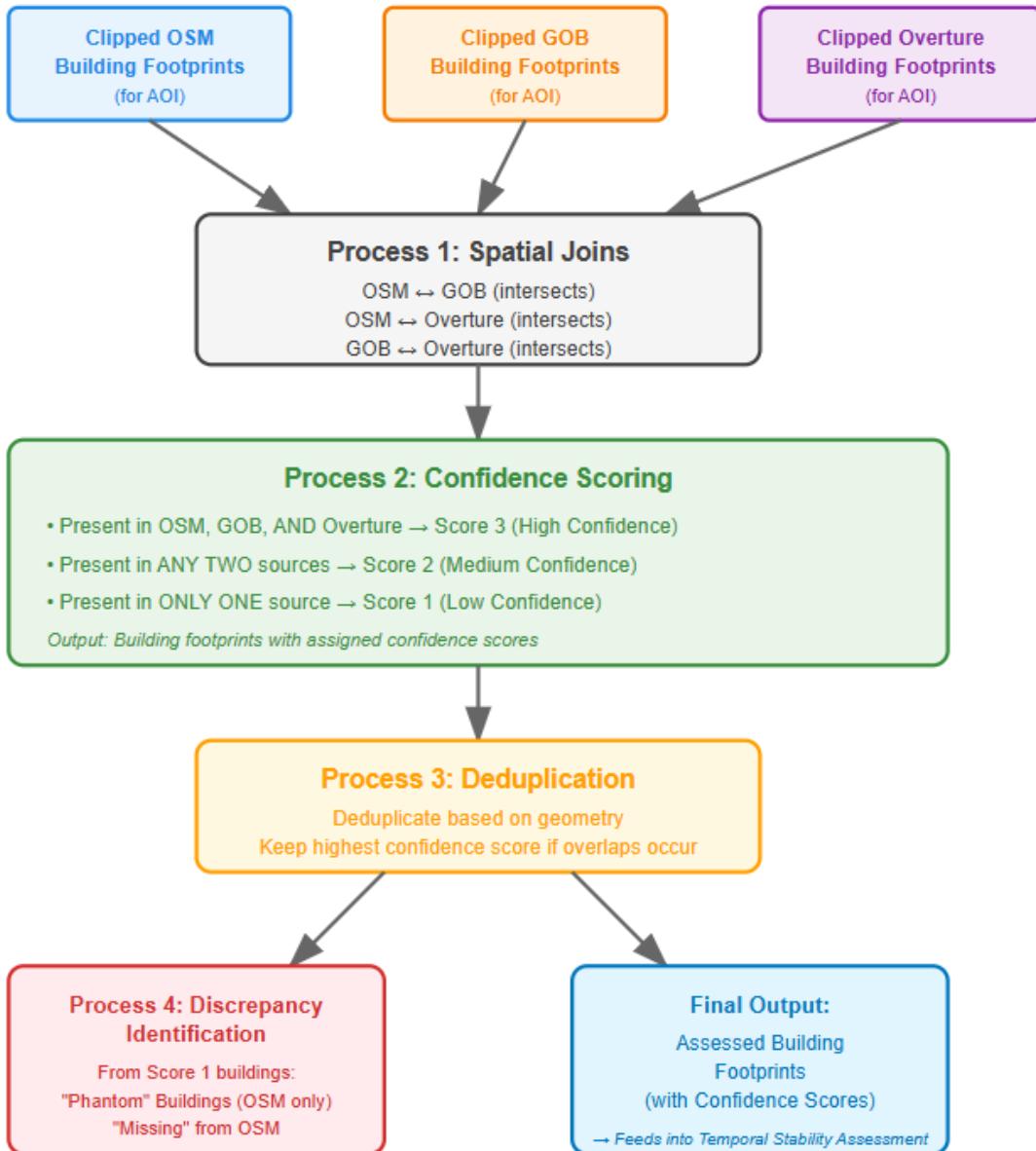


Figure 3.4: Conceptual flowchart illustrating the multi-source cross-validation process for OpenStreetMap (OSM) data quality assessment. The process involves spatial joins between OSM, Google Open Buildings (GOB), and Overture Maps building datasets, followed by confidence scoring based on inter-dataset agreement, and finally deduplication to produce a unique set of assessed building footprints. (Source: Author)

Temporal Data for Stability Assessment

To further refine the quality of the building footprints, particularly those with lower confidence (Score 1) from the multi-source assessment, a temporal stability analysis was performed. This step aimed to distinguish stable, non-vegetated built-up areas from vegetated areas or areas that underwent significant change between two observation periods, which could be misidentified as buildings or

indicate outdated footprint data.

- **Sentinel-2 Imagery for Temporal Analysis:** Two Sentinel-2 Level-2A images were selected for this analysis, capturing different seasonal vegetation states:

- Date 1: February 17, 2025 (representing a period with potentially lower vegetation activity for some species).
- Date 2: July 7, 2024 (representing a period with potentially higher vegetation activity).

These dates were chosen to maximize the potential contrast in vegetation signals. Both images were preprocessed (cloud masked, clipped to AOI) as described in Section [3.2.2](#).

- **NDVI Calculation:** The Normalized Difference Vegetation Index (NDVI) was calculated for both dates using the clipped Red (B04) and Near-Infrared (NIR, B08) bands.

- **Zonal Statistics:** For each building footprint identified with a confidence score of 1 in the multi-source assessment, the mean NDVI value was extracted from both NDVI rasters (Date 1 and Date 2) using zonal statistics (`rasterstats.zonal_stats`).

- **Temporal Stability Categorization:** Buildings were then categorized based on their temporal NDVI signatures using predefined thresholds:

- Let $\text{ndvi_diff} = \text{mean_ndvi_date2} - \text{mean_ndvi_date1}$.
- Thresholds were established: $\text{ndvi_stable_low_thresh} = 0.12$ (upper limit for non-vegetated surfaces) $\text{ndvi_stable_high_thresh} = 0.30$ (lower limit for consistently vegetated surfaces) $\text{ndvi_abs_diff_thresh} = 0.20$ (threshold for significant NDVI change)
- Categories assigned included:
 - * **Stable_NonVegetated:** Both NDVI values below `ndvi_stable_low_thresh` and $|\text{ndvi_diff}| < \text{ndvi_abs_diff_thresh}$. These are considered highly likely to be stable built-up areas.
 - * **Stable_Vegetated:** Both NDVI values above `ndvi_stable_high_thresh` and $|\text{ndvi_diff}| < \text{ndvi_abs_diff_thresh}$. These indicate stable vegetation and are likely false positives if initially tagged as buildings.

- * **Unstable_NDVI_Increase/Decrease**: $|ndvi_diff| \geq ndvi_abs_diff_thresh$.
Indicates significant change (e.g., new construction, demolition, seasonal vegetation dynamics).
 - * **Stable_IntermediateNDVI**: NDVI values are intermediate, but $|ndvi_diff| < ndvi_abs_diff_thresh$. Indicates stable but potentially mixed or non-building surfaces.
 - * **NoData_NDVI_for_Temporal**: Cases where NDVI could not be calculated (e.g., due to mask extent or no valid pixels).
- **Final Trustworthiness Assignment for Score 1 Buildings**: This temporal analysis refined the trustworthiness of the confidence score 1 buildings. For instance, Score 1 buildings categorized as **Stable_NonVegetated** (3,921 out of 149,286 initially) were promoted to a higher trustworthiness category (**Medium_Score1_StableNonVeg**), while those categorized as **Stable_Vegetated** (9 buildings) were flagged for exclusion (**Exclude_LikelyVegetation**). This comprehensive approach combining multi-source spatial agreement with temporal stability significantly reduces label noise for GNN training.

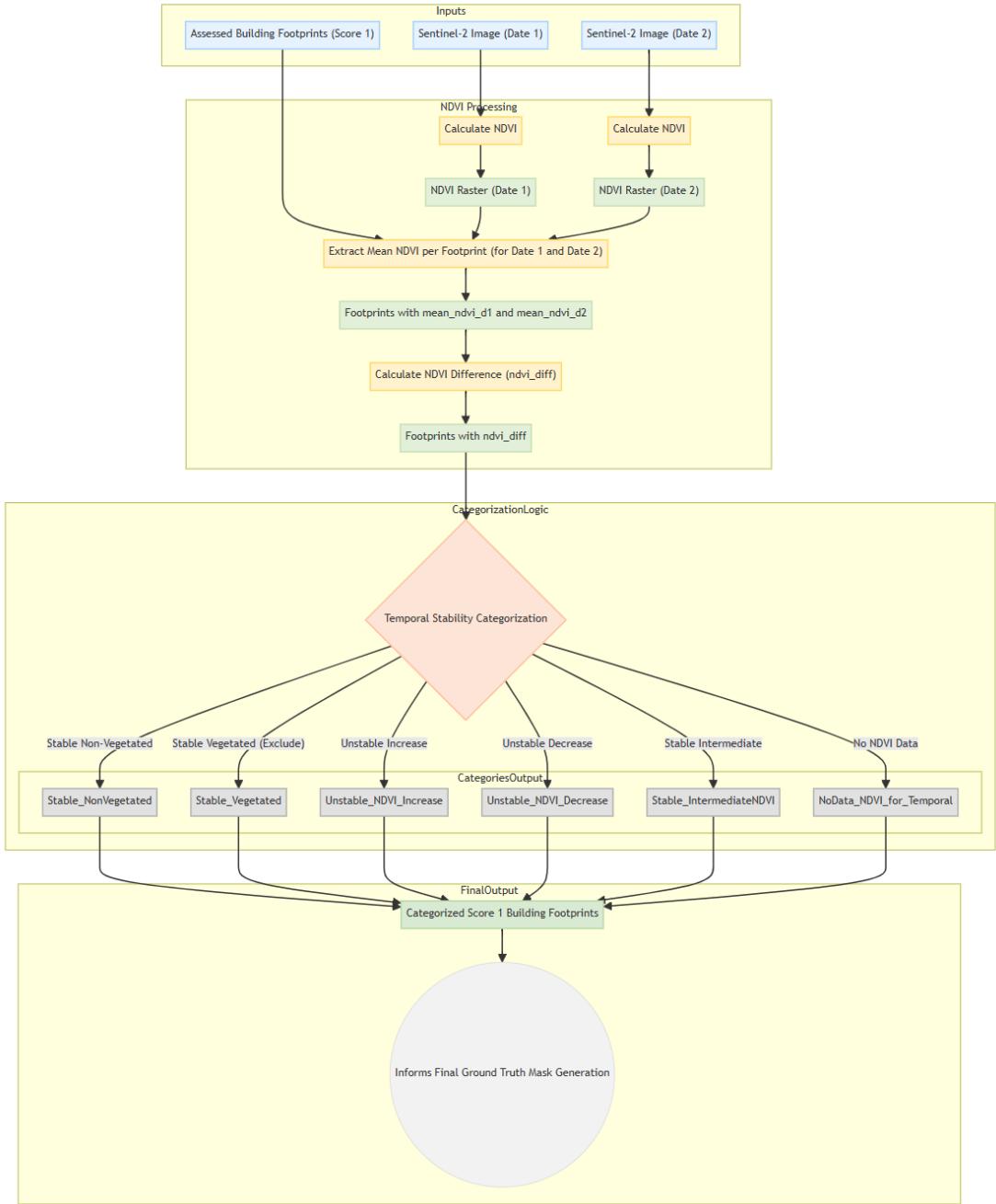


Figure 3.5: Conceptual diagram showing the logic for categorizing building footprints based on temporal NDVI analysis. This includes NDVI calculation from multi-date Sentinel-2 imagery, extraction of zonal statistics for Confidence Score 1 building footprints, NDVI difference calculation, and subsequent threshold-based categorization to assess temporal stability. (Source: Author)

Final Ground Truth Mask Generation for GNN Training

Following the multi-source cross-validation and temporal stability assessment, a final set of reliable building footprints was compiled to serve as the definitive ground truth for training the Graph Neural Network (GNN) models. This process involved selecting footprints based on their assessed confidence and temporal

stability.

The selection criteria for the final reliable building footprints were as follows:

- All building footprints with a multi-source confidence score of 2 or 3 (categorized as `High_MultiSourceAgreement`, totaling 84,717 buildings) were included. These represent buildings with strong agreement across multiple independent datasets.
- From the confidence score 1 buildings, only those that demonstrated high temporal stability characteristic of non-vegetated surfaces (categorized as `Medium_Score1_StableNonVeg` after the temporal NDVI analysis, totaling 3,921 buildings) were included.

Building footprints that were flagged during the temporal stability assessment as `Exclude_LikelyVegetation`, or those falling into `Unstable_NDVI_Increase/Decrease`, `Stable_IntermediateNDVI`, or `NoData_NDVI_for_Temporal` categories from the score 1 pool, were explicitly excluded from this final ground truth set. This rigorous selection process resulted in a combined set of 88,638 unique, reliable building footprints, which were saved as a GeoJSON file named `buildings_for_gnn_groundtruth.geojson` within the processed data directory.

These selected vector building footprints were then rasterized to create a binary building mask. This rasterization was performed to precisely align with the grid of the preprocessed Sentinel-2 imagery (10-meter resolution, same extent, transform, and Coordinate Reference System - EPSG:32631 - as the imagery processed in Section 3.2.2). The `rasterio.features.rasterize` function was employed for this task. A value of 1 was burned into the raster for pixels covered by the reliable building footprints, representing 'building' class, while a value of 0 was assigned to all other pixels, representing 'non-building' class. Crucially, the `all_touched=True` parameter was utilized during rasterization. This ensures that any pixel touched by a building polygon, even if only partially, is marked as belonging to the 'building' class. This approach is particularly important for capturing the full extent of irregularly shaped or small building features and aligns well with the superpixel-based methodology, where superpixels themselves might partially overlap true building boundaries. The resulting binary raster, saved as `aligned_reliable_urban_mask.tif` (and referred to as `final_building_mask_for_labels` in the GNN pipeline), served as the definitive ground truth for assigning labels to the superpixels during the graph construction phase (detailed in Section 3.3.2) and for the subsequent training and evaluation of all GNN models.

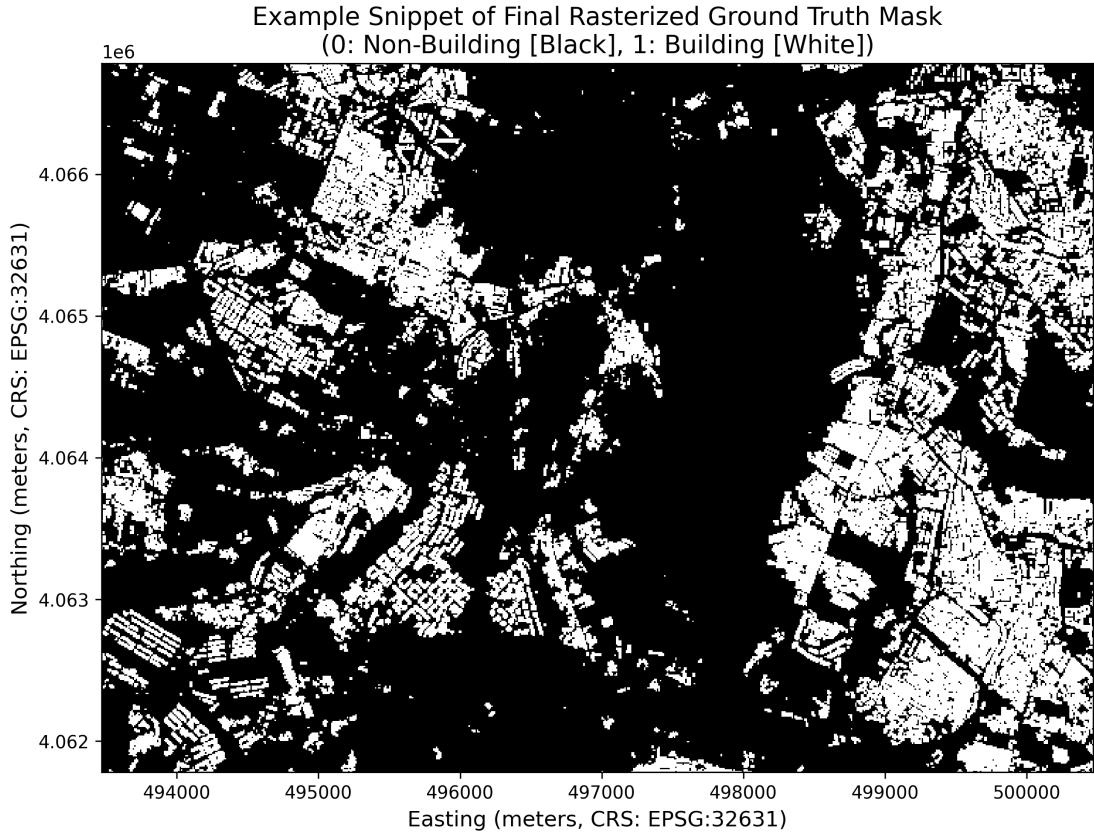


Figure 3.6: Visual example of a portion of the final binary ground truth mask (`aligned_reliable_urban_mask.tif`) for the Algiers AOI. Pixels representing building footprints (value 1) are shown in white, contrasted with non-building areas (value 0) in black. This raster mask served as the definitive ground truth for labeling superpixels. (Source: Author, derived from assessed OSM, GOB, and Overture data)

3.3 Superpixel Segmentation and Graph Construction

This section details the transformation of the preprocessed Sentinel-2 satellite imagery into a graph-based representation, which forms the fundamental input structure for the Graph Neural Network (GNN) models. This process involves two primary stages: (1) segmenting the image into perceptually meaningful regions called superpixels using the Simple Linear Iterative Clustering (SLIC) algorithm, and (2) constructing a graph where these superpixels are nodes, attributed with comprehensive spectral and geometric features, and connected by edges representing their spatial adjacency.

3.3.1 Superpixel Segmentation using SLIC

Superpixel segmentation was employed as a critical preprocessing step to group spatially contiguous pixels with similar visual characteristics into larger, more semantically meaningful "atomic regions." This approach reduces computational complexity compared to pixel-wise graph construction and provides more robust features for GNN processing by averaging out noise and local variations within each superpixel.

Simple Linear Iterative Clustering (SLIC) was selected as the superpixel generation algorithm for this research. The choice of SLIC was informed by a systematic parameter experimentation process and its known advantages, including computational efficiency, the ability to generate relatively compact and uniform superpixels, and good adherence to actual image boundaries. Unlike rigid grid-based segmentation, SLIC produces irregularly shaped superpixels that can more naturally align with the often complex and non-grid-aligned boundaries of urban features, which is crucial for accurate building footprint delineation.

The input to the SLIC algorithm was the 15-channel preprocessed and normalized Sentinel-2 composite image described in Section 3.2.2. The multi-channel nature of this input allows SLIC to leverage the rich spectral and biophysical context when grouping pixels into superpixels, going beyond simple color similarity.

A systematic experimentation process was conducted to determine the optimal SLIC parameters for the specific characteristics of the Algiers Sentinel-2 imagery, focusing on achieving a balance between granularity, boundary adherence, and internal homogeneity suitable for subsequent GNN node creation. The key parameters varied and assessed through visual analysis of full-image views and randomly selected zoomed-in regions were:

- **n_segments**: The desired number of superpixels.
- **compactness**: A parameter balancing color similarity and spatial proximity.
- **sigma**: The width of the Gaussian smoothing kernel applied before clustering.

Based on this comprehensive visual analysis and the need to balance fine-grained detail with computational efficiency for the GNN pipeline, the following parameters were selected and consistently used for segmenting all image versions (original and spectrally augmented) in this study:

- **n_segments = 20,000**
- **compactness = 1**

- `sigma = 1`
- `enforce_connectivity = True` (ensuring superpixels are contiguous)
- `channel_axis = -1` (for multi-channel imagery)

This combination was chosen as it provided the best trade-off for the task. The `n_segments` value of 20,000 offered sufficient detail to capture individual building-like structures without excessive fragmentation, which would otherwise significantly increase the number of nodes and computational burden for the GNN. For the original composite image of the Algiers AOI, this parameterization resulted in 17,916 actual unique superpixels. A `compactness` of 1 struck a good balance, allowing superpixels to adhere reasonably well to object boundaries while maintaining a somewhat regular shape suitable for node representation. A `sigma` value of 1 provided a slight smoothing effect, reducing minor noise within superpixels and promoting homogeneity without significantly blurring critical edges. The output of the SLIC segmentation process is a 2D integer array, `segments_slic`, where each pixel in the input image is assigned a unique superpixel ID (label). This array forms the direct basis for graph construction.

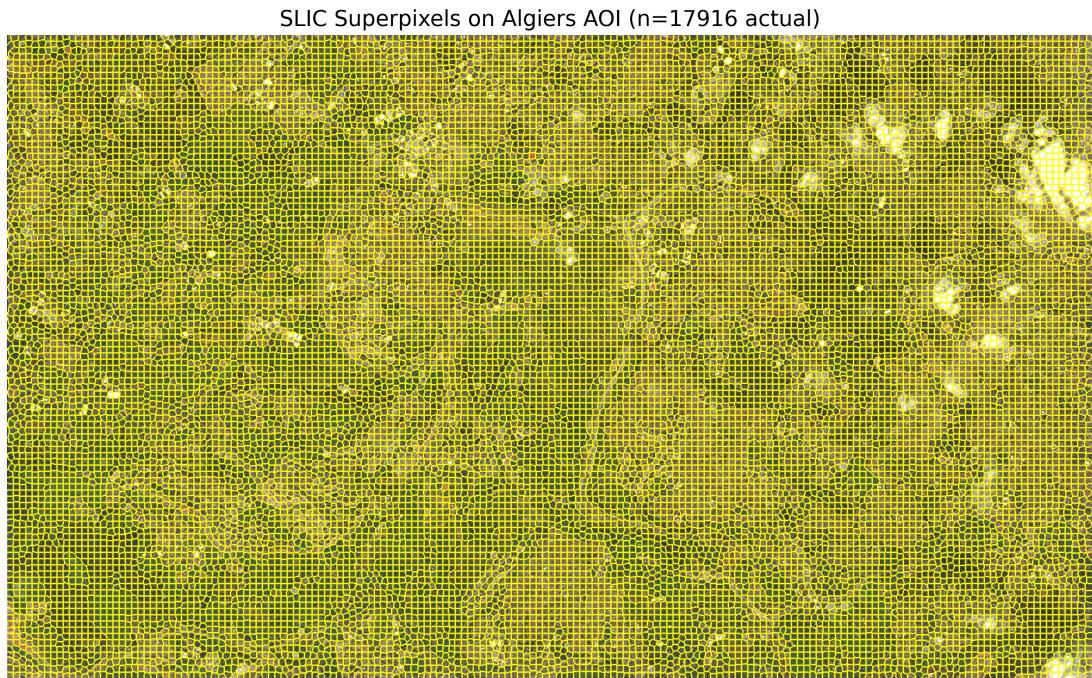


Figure 3.7: Visual example illustrating the SLIC superpixel segmentation (boundaries shown in yellow) overlaid on a Sentinel-2 RGB composite for a portion of the Algiers study area. This demonstrates the generated superpixel shapes using the chosen parameters (requested `n_segments=20,000`, `compactness=1`, `sigma=1`), resulting in 17,916 actual superpixels for the original composite image. (Source: Author, derived from Sentinel-2 imagery)

3.3.2 Graph Construction from Superpixels

Following superpixel segmentation, the image was transformed into a graph structure $G = (V, E)$, where V is the set of nodes and E is the set of edges, suitable for processing with GNNs.

- **Nodes (V):** Each unique superpixel generated by the SLIC algorithm (i.e., each unique ID in the `segments_slic` array) was treated as a distinct node in the graph. For the original composite image, this resulted in 17,916 nodes.
- **Node Features (X):** For each superpixel node, a comprehensive 19-dimensional feature vector was extracted, designed to capture its salient spectral and geometric characteristics. This feature set comprises:
 - **15 Spectral Features:** The mean pixel value for each of the 12 preprocessed Sentinel-2 bands (B01-B12, including B8A, all normalized and resampled to 10m) and the 3 derived spectral indices (NDVI, NDBI, NDSI, all normalized to [0,1]) were calculated for all pixels belonging to that superpixel. These represent the average spectral signature of each superpixel.
 - **4 Geometric Features:**
 - * `Geo_Area`: The total number of pixels comprising the superpixel.
 - * `Geo_BBoxHeight`: The height of the minimum bounding box enclosing the superpixel.
 - * `Geo_BBoxWidth`: The width of the minimum bounding box enclosing the superpixel.
 - * `Geo_AspectRatio`: The ratio of the bounding box width to its height.

These geometric features were normalized using `sklearn.preprocessing.MinMaxScaler` to scale them to a [0, 1] range based on their minimum and maximum values across all superpixels in the image. This prevents features with larger absolute values from dominating the learning process.

The resulting node feature matrix X has dimensions $N \times F$, where N is the number of superpixels (nodes) and $F = 19$ is the number of features per node.

- **Edges (E):** Edges in the graph were constructed based on the spatial adjacency of superpixels. A 4-connectivity rule was applied: an undirected

edge was established between any two superpixels if they shared a common boundary (i.e., were horizontally or vertically adjacent in the pixel grid). This process, implemented via the `create_adjacency_list` function (which iterates through pixels and their neighbors in the `segments_slic` array), results in a sparse adjacency list representing the local neighborhood relationships crucial for GNN message passing. For the original composite image with 17,916 nodes, this resulted in 45,993 unique undirected edges. The edge list was converted into a 2D NumPy array of shape (2, num_edges), representing (source_node_id, target_node_id) pairs, suitable for PyTorch Geometric. Note that for GNN processing, superpixel IDs (originally 1 to N) are typically re-indexed to be 0 to N-1.

- **Node Labels (Y) (for training and evaluation):** Each superpixel node was assigned a binary class label ('building' (1) or 'non-building' (0)). This labeling was performed by spatially overlaying the superpixel segmentation map (`segments_slic`) with the final reliable ground truth building mask (`final_building_mask_for_labels`, as generated in Section 3.2.3). A superpixel was labeled as 'building' if more than 50% of its constituent pixels overlapped with areas marked as 'building' in the ground truth mask; otherwise, it was labeled as 'non-building'. This majority voting approach ensures that the label assigned to a superpixel node accurately reflects the predominant land cover type within its boundaries. For the original composite image, this resulted in a label distribution of 12,721 non-building nodes and 5,195 building nodes, indicating a class imbalance that was considered during model training (e.g., via the Focal Loss component of the loss function).

Finally, the extracted node features (X), the constructed edge index (E), and the assigned node labels (Y) for each processed image (original and its spectrally augmented versions) were assembled into `torch_geometric.data.Data` objects. This encapsulated the graph structure and its attributes, making it directly compatible with the PyTorch Geometric library for GNN model development and training.

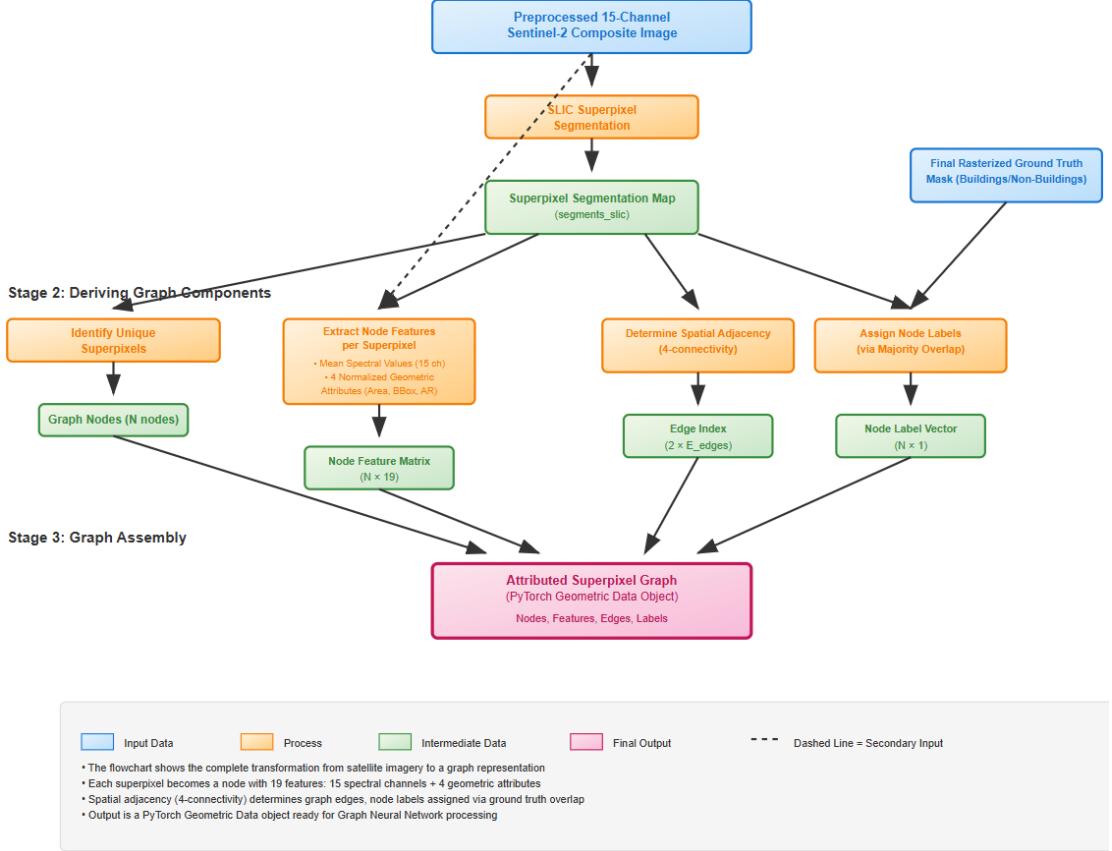


Figure 3.8: Conceptual diagram illustrating the graph construction process. Stage 1 involves SLIC superpixel segmentation of the preprocessed Sentinel-2 composite image. Stage 2 details the derivation of graph components: identifying unique superpixels as nodes, extracting a 19-dimensional feature vector (15 spectral, 4 geometric) for each node, determining spatial adjacency (4-connectivity) to form the edge index, and assigning binary node labels based on majority overlap with the final ground truth building mask. Stage 3 shows the assembly of these components into the attributed superpixel graph, a PyTorch Geometric Data object ready for GNN processing. (Source: Author)

3.4 AI/Deep Learning Model Development: UrbanGraphSAGE

This section elaborates on the architecture of the proposed Graph Neural Network (GNN) model, named UrbanGraphSAGE, specifically designed and configured for urban building footprint identification from the graph-structured superpixel data. It also details the comprehensive methodology employed for its training, including data augmentation strategies and hyperparameter choices.

3.4.1 Model Architecture: UrbanGraphSAGE

Graph Neural Networks (GNNs) were chosen for this task due to their inherent capability to process graph-structured data, effectively leveraging both node features and the topological relationships between nodes (superpixels in this context). Specifically, the GraphSAGE (Graph Sample and Aggregate) architecture **hamilton2017inductive** was selected as the foundation for the proposed model. GraphSAGE is well-suited for this application due to its inductive learning capabilities, meaning it can generalize to unseen nodes or even entirely new graphs, which is beneficial for scalability. It operates by learning functions that aggregate feature information from a node’s local neighborhood. For this research, the `mean` aggregator was chosen for its simplicity, computational efficiency, and demonstrated effectiveness in previous studies, computing the element-wise mean of neighbor embeddings **hamilton2017inductive**.

The UrbanGraphSAGE model is a multi-layer GNN designed for the binary classification of superpixel nodes into ‘building’ or ‘non-building’ categories. Each node, representing a superpixel, is characterized by the 19-dimensional feature vector detailed in Section 3.3.2. The architecture is summarized in Table 3.2 and detailed below:

The model comprises the following core components:

- **SAGEConv Layers:** The model employs a stack of three SAGEConv layers **hamilton2017inductive**. The first layer transforms the 19-dimensional input features into a 64-dimensional hidden representation. The subsequent two SAGEConv layers maintain this 64-dimensional representation. All layers utilize the ‘mean’ aggregator.
- **Batch Normalization (BatchNorm):** A BatchNorm layer follows each SAGEConv layer. This helps to stabilize the training process, accelerate convergence, and acts as a regularizer by normalizing the inputs to each layer.
- **Activation Function (ReLU):** The Rectified Linear Unit (ReLU) activation function is applied after each BatchNorm layer to introduce non-linearity, enabling the model to learn more complex patterns and relationships within the graph data [10].
- **Dropout:** A Dropout layer with a rate of 0.5 is applied after the ReLU activation in each SAGEConv layer during the training phase. This regularization technique randomly zeroes out a fraction of the input units, preventing co-adaptation of neurons and improving the model’s generalization to unseen data.

Table 3.2: UrbanGraphSAGE Model Architecture Summary

Component	Layer Type	Input Dim.	Output Dim.	Key Parameters	Activation
Input Processing	Input Features	19	19	-	-
Layer 1					
SAGEConv 1	SAGEConv	19	64	aggr='mean'	-
BatchNorm 1	BatchNorm	64	64	-	-
Activation 1	ReLU	64	64	-	ReLU
Dropout 1	Dropout	64	64	p=0.5	-
Layer 2					
SAGEConv 2	SAGEConv	64	64	aggr='mean'	-
BatchNorm 2	BatchNorm	64	64	-	-
Activation 2	ReLU	64	64	-	ReLU
Dropout 2	Dropout	64	64	p=0.5	-
Layer 3					
SAGEConv 3	SAGEConv	64	64	aggr='mean'	-
BatchNorm 3	BatchNorm	64	64	-	-
Activation 3	ReLU	64	64	-	ReLU
Dropout 3	Dropout	64	64	p=0.5	-
Feature Aggregation	JumpingKnowledge	3 * 64	192	mode='cat'	-
Classification	Linear Classifier	192	1	-	Sigmoid

The model consists of 3 SAGEConv layers, each followed by Batch Normalization, ReLU activation, and Dropout. A JumpingKnowledge layer concatenates outputs before a final linear classifier with Sigmoid activation.

- **JumpingKnowledge (JK) Layer:** A JumpingKnowledge layer with the ‘cat’ (concatenation) mode is utilized after the three SAGEConv blocks [43]. This layer aggregates the outputs from all preceding GraphSAGE layers (in this case, all 3 layers) for each node. By concatenating these representations (3 layers * 64 hidden channels/layer = 192-dimensional feature vector), the model can leverage features learned at different depths and neighborhood scopes. This allows the final classifier to access a richer, multi-scale node representation, balancing local fine-grained details with broader contextual information, which is crucial for mitigating over-smoothing in deeper GNNs.
- **Classifier:** A final fully connected linear layer (`nn.Linear(192, 1)`) acts as the classifier. It takes the 192-dimensional node representations from the JumpingKnowledge layer and maps them to a single logit.
- **Output Activation:** A Sigmoid activation function is applied to the output of the linear classifier to produce a probability score between 0 and 1, indicating the likelihood of each superpixel node belonging to the ‘building’ class.

The UrbanGraphSAGE model, as configured, has a total of 19,585 trainable parameters. This relatively compact size, given the depth and use of JumpingKnowledge, is intended to balance model capacity with the risk of overfitting, especially considering the dataset size.

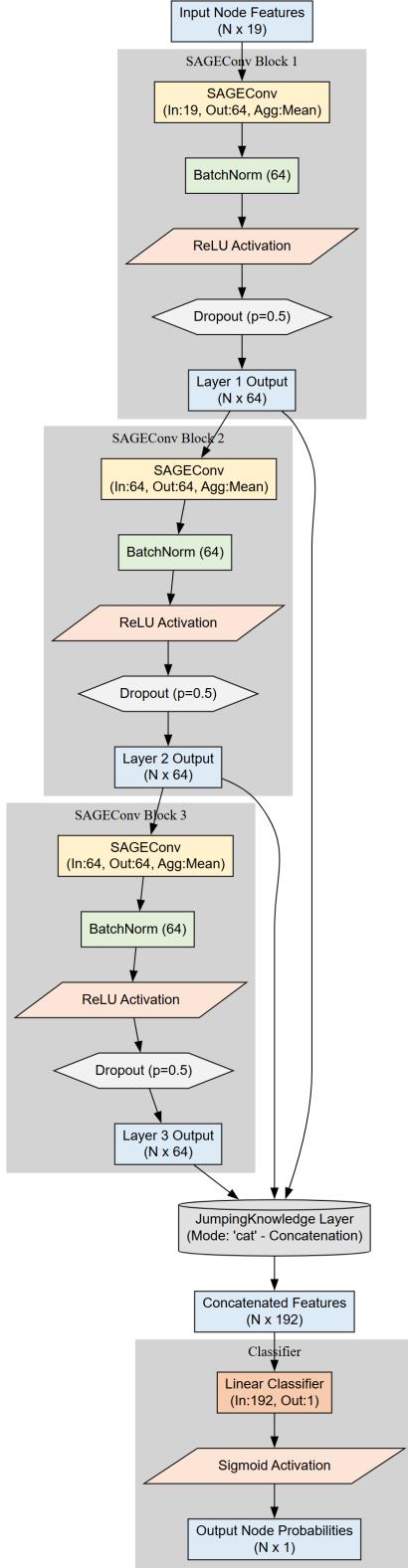


Figure 3.9: Architecture of the UrbanGraphSAGE model. The model takes $N \times 19$ node features as input and processes them through three SAGEConv blocks (each comprising SAGEConv, BatchNorm, ReLU, and Dropout). Outputs from all SAGEConv blocks are concatenated by a JumpingKnowledge ('cat' mode) layer. The resulting $N \times 192$ features are then passed to a final linear classifier with a Sigmoid activation function to produce $N \times 1$ node classification probabilities. (Source: Author)

3.4.2 Training Methodology for UrbanGraphSAGE

The training process for the UrbanGraphSAGE model was designed to leverage the prepared graph data effectively and incorporate data augmentation to enhance robustness and generalization.

- **Data Split and Masking Strategy:** A fixed node-level stratified split was applied to the graph derived from the original, non-augmented Sentinel-2 composite image. This resulted in training, validation, and test masks, with 70% of nodes allocated for training (12,541 nodes), 15% for validation (2,687 nodes), and 15% for testing (2,688 nodes). Stratification based on node labels (building/non-building) ensured proportional representation of classes in each set. These fixed masks (`train_mask`, `val_mask`, `test_mask`) were consistently used throughout all experiments, including when training with augmented data, to ensure fair and comparable evaluations.
- **Image-Level Spectral Data Augmentation:** To improve model robustness against variations in atmospheric conditions, illumination, and sensor noise, image-level spectral data augmentation was implemented. For each training epoch, in addition to the original Sentinel-2 composite image, two spectrally augmented versions were generated stochastically using `torchvision.transforms.v2`. The augmentations included:
 - Random Gaussian Blur: Kernel size (3,3), sigma uniformly sampled from (0.1, 0.5), applied with probability $p = 0.5$. This simulates minor atmospheric haze or sensor defocusing.
 - Random Brightness Adjustment (Per-Channel): Pixel values of each channel multiplied by a random factor chosen uniformly from [0.95, 1.05], applied with probability $p = 0.4$. This simulates slight variations in scene illumination or sensor response.
 - Random Contrast Adjustment (Per-Channel, Mean-Centered & Clipped): Contrast of each channel adjusted by scaling pixel values relative to the channel mean by a random factor from [0.95, 1.05], then adding the mean back and clipping to [0,1], applied with probability $p = 0.4$. This simulates variations in atmospheric clarity or sensor dynamic range.

Critically, for each of these spectrally augmented image versions, the entire preprocessing pipeline—SLIC superpixel segmentation, node feature extraction, and graph construction (adjacency list)—was re-performed. This ensured that the superpixels and their features accurately reflected the augmented image characteristics. This process resulted in a list of PyTorch

Geometric Data objects (`pyg_data_list`), typically containing the original graph and two augmented graph instances for each epoch. Graphs with node counts not matching the original graph (due to minor variations in SLIC output on augmented data, though ideally minimal with fixed `n_segments`) were filtered out to maintain compatibility with the fixed training/validation/test masks. The training loop then iterated through these graph instances, exposing the model to diverse spectral inputs while computing loss only on the nodes defined by the fixed `train_mask` derived from the original graph.

- **Optimizer:** The AdamW optimizer [44] was employed with a learning rate of 2×10^{-4} and a weight decay of 1×10^{-5} . AdamW was chosen for its effectiveness in training deep neural networks by incorporating decoupled weight decay.
- **Loss Function:** A custom `CombinedFocalDiceLoss` was utilized. This hybrid loss function combines Focal Loss [37] and Dice Loss [31].
 - Focal Loss: Addresses class imbalance (prevalence of non-building nodes) by down-weighting the loss contribution from well-classified examples and focusing training on hard-to-classify examples. Parameters were set to $\alpha \approx 0.71$ (calculated from the dataset's class distribution, giving more weight to the minority 'building' class) and $\gamma = 2.0$.
 - Dice Loss: Directly optimizes for segmentation overlap (IoU), which is highly relevant for footprint extraction tasks and effective for imbalanced segmentation. The smooth parameter was set to 1.0 to prevent division by zero.

The combined loss was weighted as $0.7 \times \text{FocalLoss} + 0.3 \times \text{DiceLoss}$. This weighting prioritizes the handling of class imbalance and hard examples via Focal Loss while still ensuring good segmentation overlap through Dice Loss.

- **Training Epochs:** The model was trained for 150 epochs.
- **Training Device:** Training was performed on a GPU (CUDA enabled, if available) to accelerate computations, managed via PyTorch's device handling.

3.4.3 Feature Importance and Selection Considerations

A permutation feature importance study was previously conducted on an earlier iteration of the UrbanGraphSAGE model, which was trained *without* the spectral data augmentation techniques described above. This study provided valuable

initial insights into the relative contributions of the 19 input features. The key findings from that prior analysis were:

- The top 5 most impactful features were identified as **B12 (SWIR 2)**, **B8A (Narrow NIR)**, **B07 (Red Edge 3)**, **B05 (Red Edge 1)**, and **B11 (SWIR 1)**.
- Reducing the feature set from 19 to 17 (by removing features with negative permutation importance on the validation set, such as Geo_Area and B01 in that study) resulted in only a minor F1-score degradation (approximately 1.3%).
- More significant reductions in feature count led to more substantial performance drops.

Table 3.3 summarizes the feature importance ranking from that study on the non-augmented model.

Table 3.3: Permutation Feature Importance Ranking for the Non-Augmented UrbanGraphSAGE Model (Validation F1 Baseline: 0.7350)

Rank	Feature Name	Drop in Validation F1-Score
1	B12 (SWIR 2)	0.0547
2	B8A (Narrow NIR)	0.0157
3	B07 (Red Edge 3)	0.0128
4	B05 (Red Edge 1)	0.0120
5	B11 (SWIR 1)	0.0120
6	B04 (Red)	0.0096
7	B08 (NIR)	0.0083
8	B09 (Water Vapour)	0.0073
9	B03 (Green)	0.0060
10	B02 (Blue)	0.0052
11	NDBI	0.0040
12	NDVI	0.0036
13	Geo_Area	0.0018
14	Geo_BBoxWidth	0.0003
15	Geo_AspectRatio	-0.0011
16	B01 (Coastal Aerosol)	-0.0015
17	B06 (Red Edge 2)	-0.0016
18	Geo_BBoxHeight	-0.0021
19	NDSI	-0.0064

Features are ranked from most to least important based on the drop in validation F1 score when permuted. Negative values indicate permutation improved performance, suggesting the feature might be noisy or redundant for that specific non-augmented model context.

Decision for the Current Augmented Model: Despite the insights from the prior feature selection study on the non-augmented model, the final Urban-GraphSAGE model detailed in this thesis (trained **with** spectral data augmentation) utilized the ***full 19-feature set***. This decision was primarily driven by empirical results from the complete training pipeline, where the model employing all 19 features alongside spectral augmentation achieved the highest overall performance (Test F1 Score: 0.7579). It was hypothesized that spectral augmentation might alter feature interactions and importance, and that some features deemed less critical or even detrimental in the non-augmented context could contribute positively when the model is trained on more diverse data. Re-running the entire permutation feature importance analysis and subsequent retraining with reduced feature sets on the **augmented** model was considered outside the scope of the immediate research timeline due to computational constraints. Therefore, to maximize performance with the developed augmented training approach, the complete 19-feature set was retained. The prior feature importance study, however, still provides valuable general insights into the utility of different feature types (spectral bands vs. indices vs. geometric features) for this GNN architecture in the context of building footprint extraction.

3.5 Implementation and Evaluation

This section outlines the metrics employed to assess the performance of the trained Graph Neural Network (GNN) models, the specific evaluation protocol followed, any baselines used for comparative analysis, and the computational environment in which the research was conducted.

3.5.1 Evaluation Metrics

The performance of the trained GNN models in identifying urban building footprints (a binary classification task at the superpixel node level) was quantitatively assessed using a suite of standard classification and segmentation metrics. These metrics provide a comprehensive understanding of the model’s accuracy, its ability to correctly identify building and non-building superpixels, and its performance in handling class imbalance. The definitions for these metrics are provided in Table 3.4.

Table 3.4: Evaluation Metrics Definitions

Metric Name	Formula	Description/Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Proportion of correctly classified superpixels (buildings and non-buildings) out of the total.
Precision (for 'building' class)	$\frac{TP}{TP+FP}$	Proportion of correctly identified building superpixels among all superpixels predicted as 'building'. Indicates avoidance of false positive building predictions.
Recall (Sensitivity, for 'building' class)	$\frac{TP}{TP+FN}$	Proportion of correctly identified building superpixels among all actual building superpixels. Indicates ability to detect all relevant building instances.
F1-Score (for 'building' class)	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of Precision and Recall. Provides a balanced measure, crucial for imbalanced classification tasks.
Confusion Matrix	A 2x2 matrix showing counts of: True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN)	Provides a detailed error analysis, showing the types of errors made by the model (e.g., buildings misclassified as non-buildings, and vice-versa).

TP: True Positives (building correctly identified as building), TN: True Negatives (non-building correctly identified as non-building), FP: False Positives (non-building incorrectly identified as building), FN: False Negatives (building incorrectly identified as non-building).

The **F1-Score** for the 'building' class was chosen as the primary evaluation metric. This metric is particularly well-suited for tasks with imbalanced class distributions, such as urban area identification where non-building areas often significantly outnumber building areas. The F1-Score provides a balanced measure between Precision and Recall, penalizing models that perform poorly on the minority class (buildings) even if overall accuracy is high due to correctly classifying the majority class (non-building). While the F1-Score was primary, Accuracy, Precision, and Recall were also reported to provide a more nuanced understanding of model behavior. The Confusion Matrix was generated for each evaluation to allow for a detailed inspection of the types of classification errors made by the

models.

3.5.2 Evaluation Protocol

To ensure a fair and unbiased assessment of the generalization capabilities of the trained GNN models, the following evaluation protocol was strictly adhered to:

- **Held-out Datasets:** All models (UrbanGraphSAGE, UrbanGCN_Enhanced, UrbanGAT), whether trained with or without spectral augmentation, were evaluated on the fixed, held-out validation and test sets. These sets were derived from the graph constructed using the original, non-augmented Sentinel-2 composite image and its corresponding assessed ground truth labels (as defined in Section 3.4.2). This ensures that the evaluation is performed on data not seen during the training or augmentation generation process, providing a true measure of how well the models generalize.
- **Performance Monitoring during Training:** While the final reported metrics are from the test set, the model’s performance on the validation set was monitored periodically during the training phase (every 10 epochs). This monitoring helped in understanding the learning progression, guiding hyperparameter tuning decisions (though extensive tuning was not the primary focus of this stage for all models), and identifying potential overfitting by comparing training loss/metrics with validation loss/metrics. Explicit early stopping based on validation performance was not detailed as a formal part of the training loop in the provided materials but is a common practice this monitoring would support.
- **Comparative Analysis:** The performance of the primary model, UrbanGraphSAGE (trained with spectral augmentation), was compared against:
 - Its non-augmented counterpart (UrbanGraphSAGE trained without spectral augmentation, using the baseline F1-score from the feature importance study as a reference).
 - Other GNN architectures (UrbanGCN_Enhanced and UrbanGAT) trained with the same spectral augmentation strategy and assessed ground truth data.

This multi-model comparison helps to contextualize the performance of the UrbanGraphSAGE architecture.

3.5.3 Baselines for Comparison

The primary baseline for evaluating the UrbanGraphSAGE model trained with spectral augmentation was its own performance when trained *without* spectral augmentation. The results from the feature importance study (Section 3.4.3), particularly the validation F1-score of the non-augmented UrbanGraphSAGE model (e.g., 0.7350), served as this internal baseline to quantify the direct impact of the data augmentation strategy.

Furthermore, two additional GNN architectures, also trained with the same spectral augmentation techniques and assessed ground truth data, served as strong deep learning baselines:

- **UrbanGCN_Enhanced:** A 3-layer Graph Convolutional Network (GCN) model with skip connections and hidden channels of 128, 256, and 128 respectively.
- **UrbanGAT:** A 3-layer Graph Attention Network (GAT) model, with 8 attention heads per layer and 8 hidden channels per head in each layer (resulting in 64-dimensional outputs per GAT layer before potential concatenation or final processing).

The comparative performance of UrbanGraphSAGE against these GCN and GAT models (e.g., Test F1 Scores: UrbanGraphSAGE 0.7579, GCN 0.7535, GAT 0.6486, as per your notebook results) provides insights into the relative effectiveness of the GraphSAGE architecture for this specific task and dataset configuration.

3.5.4 Software and Hardware Environment

The entire methodology, from data preprocessing to model training and evaluation, was implemented using Python (version 3.11, as indicated by package installation paths in notebook outputs). Key software libraries and frameworks utilized include:

- **Deep Learning Frameworks:** PyTorch (version 2.6.0+cu124), PyTorch Geometric (version 2.6.1), and Torchvision (version 0.21.0+cu124).
- **Geospatial Data Processing:** Rasterio (version 1.4.3), GeoPandas (version 1.1.0), Shapely (version 2.1.1), and Pyproj (version 3.7.1).
- **Image Processing and Scientific Computing:** Scikit-image (version 0.25.2), NumPy (version 2.2.6), Pandas (version 2.3.0), Matplotlib (version 3.10.3), and Seaborn (for heatmap visualization).

- **Data Augmentation (Image-Level):** The primary spectral augmentation at the image level was implemented using the `torchvision.transforms.v2` library (version 0.21.0+cu124).

The computational environment for model training and extensive data processing was Google Colaboratory, leveraging its cloud-based resources. Training of the GNN models was performed on a GPU, specifically utilizing CUDA 12.4 (as indicated by PyTorch versioning), which significantly accelerated the deep learning computations. The specific GPU model provided by Google Colab varies but is typically an NVIDIA Tesla T4, K80, P100, or similar.

Chapter 4

Results and Discussion

This chapter presents a comprehensive overview, analysis, and interpretation of the results obtained from the application of the methodology detailed in Chapter 3. The primary objective is to evaluate the performance of the UrbanGraphSAGE model for urban building footprint identification in Algiers, assess the impact of its key methodological components (assessed OpenStreetMap data and spectral augmentation), and contextualize these findings within the broader field of remote sensing and urban planning.

4.1 Presentation of Results

This section systematically presents the quantitative and qualitative outcomes of the experimental phase. It focuses on the performance of the primary UrbanGraphSAGE model and its constituent components, alongside comparative results from other GNN architectures investigated.

4.1.1 Performance of UrbanGraphSAGE Model (with Spectral Augmentation and Assessed OSM Data)

The primary UrbanGraphSAGE model, meticulously trained with rigorously assessed OpenStreetMap (OSM) data and further enhanced by the application of image-level spectral data augmentation techniques (as detailed in Sections 3.2.3 and 3.4.2 respectively), demonstrated robust and promising performance in the task of identifying urban building footprints within the Algiers study area.

The model achieved a **Test F1-Score of 0.7579**, indicating a strong balance between precision and recall in its classification of superpixel nodes. The comprehensive performance metrics on the held-out test set are summarized in Table 4.1. The overall accuracy, representing the proportion of correctly classified superpixels (both urban and non-urban), reached 0.8296. Precision, which measures

the proportion of correctly identified urban superpixels among all superpixels predicted as urban by the model, was 0.6448. Notably, the recall, which quantifies the proportion of correctly identified urban superpixels among all actual urban superpixels, was high at 0.9192. The F1-Score, as the harmonic mean of precision and recall, provides a critical balanced measure, particularly for imbalanced classification tasks such as building extraction, where the 'urban' class is typically less prevalent than the 'non-urban' class. The test set for this evaluation comprised 2,688 total superpixel nodes, of which 780 were actual urban nodes, meaning urban superpixels constituted approximately 28.9% of the test data.

Table 4.1: Performance Metrics of the UrbanGraphSAGE Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.

Metric	Value
Accuracy	0.8296
Precision	0.6448
Recall	0.9192
F1-Score	0.7579

The detailed classification results on the test set are further elucidated by the confusion matrix, presented in Table 4.2 and visualized in Figure 4.1. The model correctly identified 717 true urban superpixels (True Positives, TP) and 1513 true non-urban superpixels (True Negatives, TN). However, it misclassified 395 non-urban superpixels as urban (False Positives, FP), representing commission errors. Conversely, it missed only 63 actual urban superpixels (False Negatives, FN), representing omission errors.

Table 4.2: Confusion Matrix for the UrbanGraphSAGE Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.

	Predicted Non-Urban	Predicted Urban
True Non-Urban	1513 (TN)	395 (FP)
True Urban	63 (FN)	717 (TP)

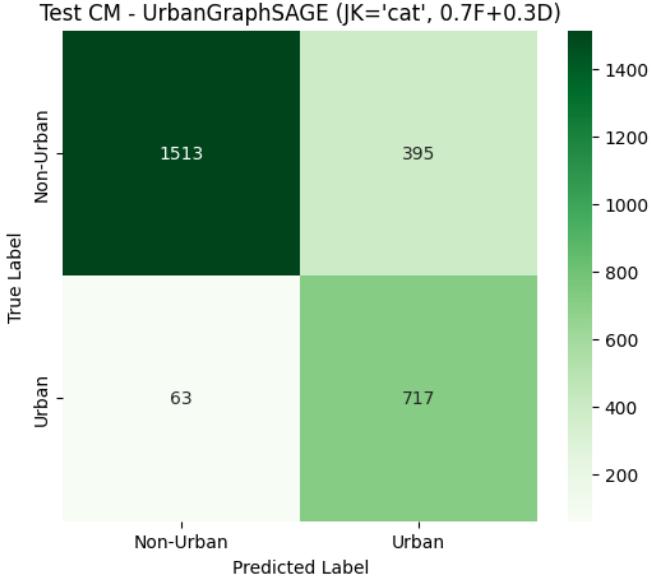


Figure 4.1: Visual representation of the confusion matrix for the UrbanGraphSAGE model (trained with spectral augmentation and assessed OSM data) on the test set. The matrix illustrates True Positives (TP=717), True Negatives (TN=1513), False Positives (FP=395), and False Negatives (FN=63). (Source: Author)

The notably high recall of 0.9192 indicates that the UrbanGraphSAGE model is highly effective in identifying the majority of true urban superpixels within the Algiers study area. This demonstrates a strong capability to detect existing building footprints, thereby minimizing omission errors, which is a critical attribute for applications requiring comprehensive mapping, such as urban growth monitoring or inventory creation for disaster management. The observed high recall, coupled with a more moderate precision, suggests a learning strategy that prioritizes completeness in detection. This behavior is significantly influenced by the chosen `CombinedFocalDiceLoss` function, which was configured with weights `weight_focal=0.7` and `weight_dice=0.3`. The Dice Loss component inherently optimizes for the overlap between predicted and true masks, which often encourages models to minimize false negatives, thus boosting recall. Furthermore, the Focal Loss component, with its `focal_alpha` parameter set to approximately 0.7100 (favoring the minority 'urban' class), further reinforces the model's attention towards correctly identifying building superpixels. This trade-off, where completeness of detection is favored, potentially at the cost of a higher rate of false alarms (commission errors), aligns with practical considerations in many remote sensing applications. For instance, in urban growth monitoring or disaster damage assessment, failing to detect an existing building (omission error) can have more significant downstream consequences than incorrectly identifying some non-building features as urban (commission error), especially if the latter can be

refined or filtered in subsequent post-processing stages.

4.1.2 Comparative Performance with GCN and GAT Models

To contextualize the performance of the primary UrbanGraphSAGE model, its results were rigorously compared against two other prominent Graph Neural Network architectures: a Graph Convolutional Network (UrbanGCN_Enhanced) and a Graph Attention Network (UrbanGAT). It is crucial to note that all comparative models were trained under identical conditions to UrbanGraphSAGE. This included the utilization of the same rigorously assessed OpenStreetMap (OSM) data for ground truth, the same image-level spectral data augmentation techniques applied during training, the same 150 training epochs, and the same `CombinedFocalDiceLoss` function with identical weighting and parameters. This consistent experimental setup ensures that observed performance differences are primarily attributable to the architectural variations between the GNN models themselves.

The UrbanGraphSAGE model demonstrated a slight numerical advantage in terms of the primary F1-Score metric over UrbanGCN_Enhanced, and a more significant lead over the UrbanGAT model. Specifically, on the test set:

- UrbanGraphSAGE achieved an F1-Score of **0.7579**.
- UrbanGCN_Enhanced recorded an F1-Score of **0.7535**.
- UrbanGAT yielded an F1-Score of **0.6486**.

Table 4.3 provides a direct comparison of these key performance indicators.

Table 4.3: Comparative Test F1-Scores of UrbanGraphSAGE, UrbanGCN_Enhanced, and UrbanGAT Models (All trained with spectral augmentation and assessed OSM data).

Model Name	Test F1-Score
UrbanGraphSAGE	0.7579
UrbanGCN_Enhanced	0.7535
UrbanGAT	0.6486

The detailed classification results for these comparative GNN models on the test set are further elucidated by their respective confusion matrices.

For the **UrbanGCN_Enhanced** model, the test confusion matrix (Table 4.4, visualized in Figure 4.2) showed 1588 True Negatives (TN), 320 False Positives

(FP), 115 False Negatives (FN), and 665 True Positives (TP). This translated to an accuracy of 0.8382, precision of 0.6751, and recall of 0.8526.

Table 4.4: Confusion Matrix for the UrbanGCN_Enhanced Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.

	Predicted Non-Urban	Predicted Urban
True Non-Urban	1588 (TN)	320 (FP)
True Urban	115 (FN)	665 (TP)

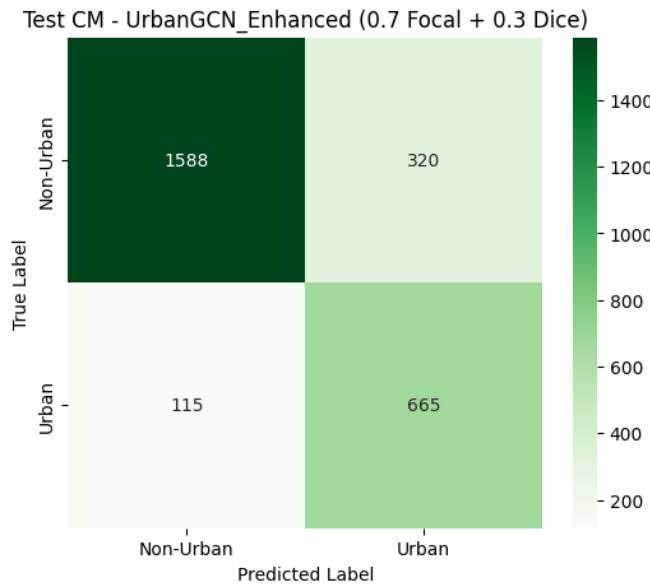


Figure 4.2: Visual representation of the confusion matrix for the UrbanGCN_Enhanced model (trained with spectral augmentation and assessed OSM data) on the test set, illustrating True Positives (TP=665), True Negatives (TN=1588), False Positives (FP=320), and False Negatives (FN=115). (Source: Author)

For the **UrbanGAT** model, the test confusion matrix (Table 4.5, visualized in Figure 4.3) indicated 1365 True Negatives (TN), 543 False Positives (FP), 145 False Negatives (FN), and 635 True Positives (TP). This resulted in an accuracy of 0.7440, precision of 0.5390, and recall of 0.8141.

Table 4.5: Confusion Matrix for the UrbanGAT Model (Trained with Spectral Augmentation and Assessed OSM Data) on the Test Set.

	Predicted Non-Urban	Predicted Urban
True Non-Urban	1365 (TN)	543 (FP)
True Urban	145 (FN)	635 (TP)

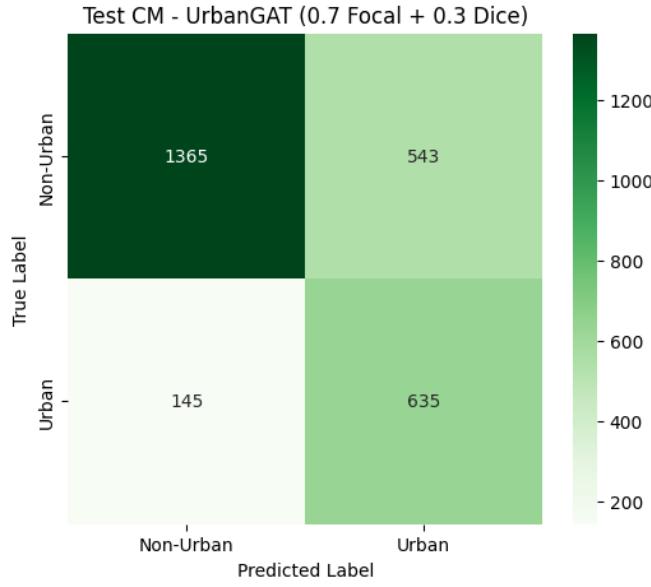


Figure 4.3: Visual representation of the confusion matrix for the UrbanGAT model (trained with spectral augmentation and assessed OSM data) on the test set, illustrating True Positives (TP=635), True Negatives (TN=1365), False Positives (FP=543), and False Negatives (FN=145). (Source: Author)

A visual comparison of the F1-Scores for all three augmented GNN models is provided in Figure 4.4.

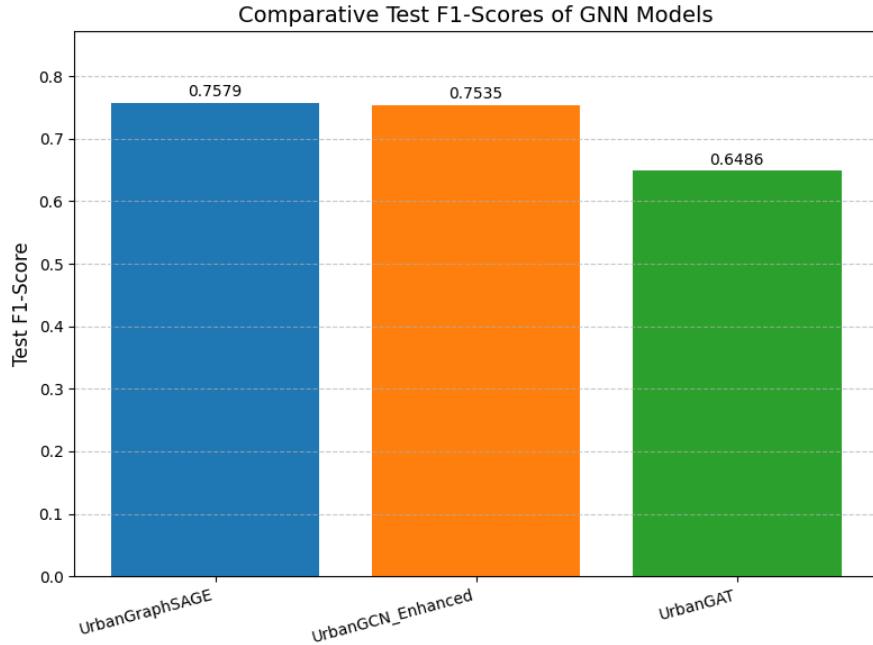


Figure 4.4: Comparison of Test F1-Scores achieved by UrbanGraphSAGE, UrbanGCN_Enhanced, and UrbanGAT models. All models were trained with spectral augmentation and assessed OpenStreetMap (OSM) data. (Source: Author)

The marginal difference in F1-Scores between UrbanGraphSAGE (0.7579) and

UrbanGCN_Enhanced (0.7535), a mere 0.0044, suggests that for this specific task and dataset configuration, the inductive learning capabilities and sampling strategy of GraphSAGE offer only a slight performance advantage over the more straightforward message passing and full neighborhood aggregation employed by GCNs. A closer look at precision and recall reveals nuanced differences: UrbanGraphSAGE achieved a higher recall (0.9192 vs. 0.8526 for GCN), while UrbanGCN_Enhanced exhibited slightly better precision (0.6751 vs. 0.6448 for UrbanGraphSAGE). This indicates that while UrbanGraphSAGE was marginally more effective at detecting a greater proportion of true building superpixels (minimizing omissions), UrbanGCN_Enhanced was slightly more adept at minimizing false alarms (incorrectly classifying non-building superpixels as buildings). The UrbanGAT model demonstrated substantially lower performance across all metrics compared to both UrbanGraphSAGE and UrbanGCN_Enhanced. Its lower precision (0.5390) and the highest number of false positives (543) suggest particular challenges with its attention mechanism in the context of the noisy and heterogeneous urban landscape of Algiers, potentially leading to over-attention to irrelevant features or difficulties in generalizing attention weights effectively.

4.1.3 Impact of Spectral Data Augmentation

Image-level spectral data augmentation was a strategic component of the methodology, designed to enhance the UrbanGraphSAGE model’s robustness to variations in input data and improve its generalization capabilities to unseen Sentinel-2 imagery. The core idea was to expose the model during training to a wider variety of plausible spectral conditions than might be present in the limited set of original training scenes.

The positive impact of this strategy is demonstrated by a direct comparison of the final test scores. The augmented UrbanGraphSAGE model achieved a **Test F1-Score of 0.7579**. For a fair comparison, this is benchmarked against the test performance of the non-augmented baseline model, which achieved a **Test F1-Score of 0.7297**. The application of spectral augmentation, therefore, resulted in a notable **F1-Score increase of 0.0282**.

This improvement underscores the value of spectral augmentation as an effective regularization technique. The augmentation pipeline (detailed in Section 3.4.2) involved applying random Gaussian blur, brightness adjustments, and contrast modifications to the 15-channel composite Sentinel-2 image. These transformations simulate common real-world variations in:

- **Atmospheric Conditions:** Gaussian blur can mimic the effects of slight atmospheric haze or aerosol scattering.

- **Illumination Changes:** Brightness adjustments simulate variations due to different sun angles, times of day, or slight sensor calibration differences.
- **Sensor Noise and Dynamic Range:** Contrast modifications can represent variations in sensor dynamic range or minor noise artifacts.

By training the UrbanGraphSAGE model on these diverse augmented versions of the input imagery (where superpixel segmentation and feature extraction were re-performed for each augmented instance), the model was compelled to learn features that are more invariant to these minor spectral shifts and radiometric distortions. This effectively increased the diversity and effective size of the training dataset without requiring the acquisition and laborious labeling of additional unique geographic scenes.

The consequence of learning these more robust and generalized features is an improved ability of the model to perform consistently on unseen test data, which may itself exhibit slightly different spectral characteristics compared to the original training images. This is particularly crucial in remote sensing applications, where acquiring a perfectly representative and exhaustive ground truth dataset covering all possible environmental and sensor-related variations is often impractical and cost-prohibitive. Spectral augmentation provides a data-efficient pathway to enhance model resilience, making the trained models more broadly applicable across different acquisition dates, subtle seasonal changes, or even imagery from similar sensors, potentially reducing the need for extensive re-training or re-calibration for new deployments. The observed quantitative improvement in the F1-Score directly validates this critical role of spectral augmentation in the proposed building footprint extraction framework.

4.1.4 Insights from SLIC Parameter Experimentation

The quality of the superpixel segmentation, which provides the foundational graph nodes for the GNN, directly influences the subsequent feature extraction and model performance. Therefore, a systematic experimentation with the key parameters of the Simple Linear Iterative Clustering (SLIC) algorithm was conducted to optimize superpixel generation for the 15-channel Sentinel-2 composite image of the Algiers study area. The findings of this experimentation were crucial in selecting parameters that offered the best trade-off between detail capture, boundary adherence, computational load, and semantic coherence for the building footprint identification task.

The experimentation involved systematically varying one SLIC parameter at a time while keeping others at defined baseline values (`n_segments_baseline =`

`10000, compactness_baseline = 1, sigma_baseline = 1`). The impact of these variations was assessed through visual analysis of both full-image segmentations and randomly selected zoomed-in regions, focusing on granularity, adherence to urban feature boundaries, and internal homogeneity of superpixels. Key observations from this process included:

- **Varying n_segments (Desired Number of Superpixels):**
 - Lower values (e.g., `n_segments=5000`, resulting in 5,058 actual superpixels) led to noticeably larger superpixels. While computationally lighter for downstream GNN processing, this often resulted in significant under-segmentation in dense urban areas, merging distinct small buildings or features, which was clearly visible in zoomed-in views.
 - Higher values (e.g., `n_segments=50000`, resulting in 58,403 actual superpixels) produced very small, highly granular superpixels. While offering fine detail, this ran a considerable risk of over-segmentation, breaking down individual buildings into too many parts and potentially capturing more noise. Furthermore, such a high number of nodes would substantially increase the computational load for graph construction and GNN training.
 - Values around the baseline of 10,000 (10,212 actual superpixels) and the chosen 20,000 (17,916 actual superpixels for the original image using final parameters) appeared to offer a more optimal balance, providing reasonable detail for urban features without excessive fragmentation or computational burden. The choice of 20,000 for the final pipeline aimed to capture smaller urban features more effectively than 10,000 segments.
- **Varying compactness (Spatial Regularity vs. Boundary Adherence):**
 - Low compactness values (e.g., 0.1 and 0.5, resulting in 7,400 and 10,069 actual superpixels respectively when `n_segments=10000` was requested) led to superpixels that adhered more closely to spectral boundaries, resulting in more irregular shapes. While `compactness=0.1` showed good outline following in some zoomed regions, it also significantly reduced the actual number of superpixels generated compared to the requested number, indicating more merging based on spectral similarity.
 - High compactness values (e.g., 10 and 30, both resulting in 10,218 actual superpixels when `n_segments=10000` was requested) produced

very regular, grid-like superpixels. These often failed to respect natural or man-made boundaries, cutting across distinct urban features, which is undesirable for accurate object-based classification.

- A **compactness** value around 1.0 (the baseline, yielding 10,212 actual superpixels with **n_segments=10000**) or 0.5 seemed to provide a good compromise, generally following object boundaries with some degree of regularity, making them suitable for node representation. The final choice of **compactness=1** for the GNN pipeline aimed for this balance.

- **Varying sigma (Smoothing):**

- A **sigma** value of 0 (no smoothing, resulting in 10,180 actual superpixels with baseline **n_segments/compactness**) led to superpixels that were highly sensitive to local variations and noise, often resulting in somewhat jagged or busy boundaries in zoomed views.
- Higher **sigma** values (e.g., 3 and 5, both resulting in 10,218 actual superpixels) produced visibly smoother and more homogeneous superpixels but also tended to blur fine edges and details, potentially merging distinct but spectrally similar close features.
- A small amount of smoothing, such as **sigma=1** (the baseline, yielding 10,212 actual superpixels), appeared beneficial for reducing minor noise and creating more homogeneous superpixels while largely preserving important feature boundaries critical for distinguishing buildings. This was the value chosen for the final GNN pipeline.

Based on this comprehensive visual analysis, the SLIC parameters selected for the main GNN processing pipeline were **n_segments = 20,000**, **compactness = 1**, and **sigma = 1**. This parameter set, resulting in 17,916 actual superpixels for the original composite image, was deemed optimal for providing a good level of detail for capturing smaller urban features (better than 10,000 segments), maintaining decent adherence to object boundaries for meaningful node representation (better than high compactness), and offering slight smoothing to reduce minor noise without significantly blurring critical edges (better than **sigma=0** or very high **sigma**). The iterative and visual-centric approach to SLIC parameter tuning was essential, as purely statistical metrics for segmentation quality do not always capture the semantic coherence required for generating effective graph nodes for GNNs. The goal was to create superpixels that accurately represent real-world building blocks for subsequent graph-based learning.

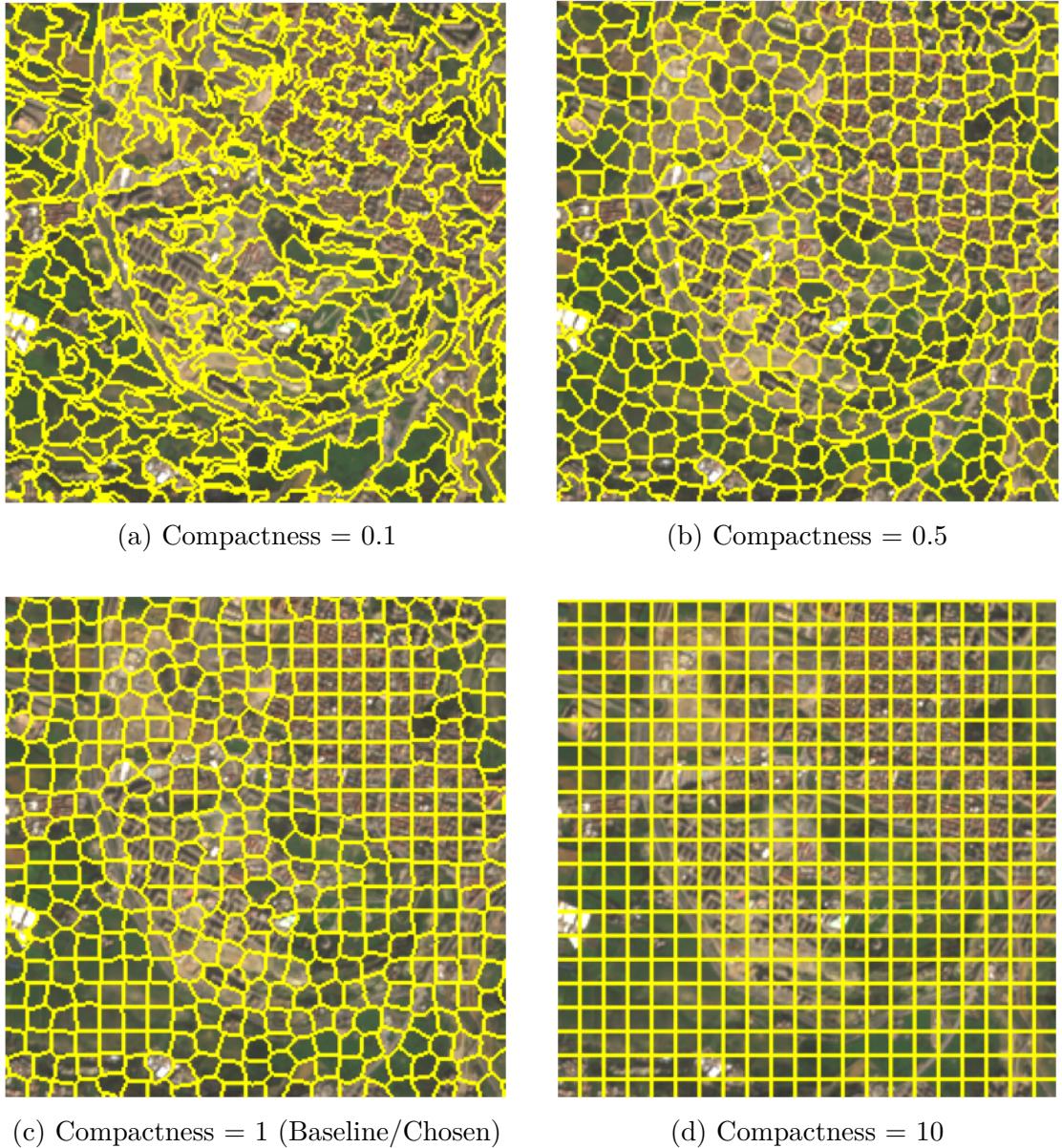


Figure 4.5: Illustrative comparison of SLIC segmentation results on a representative zoomed-in area of Algiers, showcasing the impact of varying the `compactness` parameter. All images used requested `n_segments=10,000` and `sigma=1`. (a) Compactness=0.1 shows highly irregular superpixels strongly adhering to spectral boundaries. (b) Compactness=0.5 shows a better balance. (c) Compactness=1 (chosen for final GNN pipeline along with `n_segments=20,000`) offers a good compromise between boundary adherence and regularity. (d) Compactness=10 results in very regular, grid-like superpixels often disregarding feature boundaries. (Source: Author, derived from Sentinel-2 imagery)

4.1.5 Findings from Permutation Feature Importance (on Non-Augmented Model)

To understand the relative contribution of the 19 input features (15 spectral, 4 geometric) to the predictive capability of the UrbanGraphSAGE architecture, a

permutation feature importance study was conducted. It is important to note that this particular analysis was performed on an earlier iteration of the UrbanGraphSAGE model that was trained **without** the image-level spectral data augmentation techniques described in Section 3.4.2. The baseline performance for this non-augmented model, utilizing all 19 features, was a Validation F1-Score of 0.7350 (as referenced in Section 3.4.3).

The permutation feature importance method assesses the drop in model performance (Validation F1-Score in this case) when the values of a single feature are randomly shuffled across the validation set, thereby breaking its relationship with the target variable. A larger drop indicates higher feature importance. The detailed ranking of features by their importance, based on this drop from the baseline Validation F1-Score of 0.7350, is presented in Table 4.6 (which reiterates Table 3.3 from Chapter 3 for clarity in this results section) and visualized in Figure 4.6.

Table 4.6: Permutation Feature Importance Ranking for the Non-Augmented UrbanGraphSAGE Model (Validation F1 Baseline: 0.7350). Features are ranked by the drop in Validation F1-Score when permuted.

Rank	Feature Name	Importance (Drop in Validation F1-Score)
1	B12 (SWIR 2)	0.054728
2	B8A (Narrow NIR)	0.015682
3	B07 (Red Edge 3)	0.012807
4	B05 (Red Edge 1)	0.011986
5	B11 (SWIR 1)	0.011986
6	B04 (Red)	0.009573
7	B08 (NIR)	0.008298
8	B09 (Water Vapour)	0.007331
9	B03 (Green)	0.006038
10	B02 (Blue)	0.005157
11	NDBI	0.003951
12	NDVI	0.003602
13	Geo_Area	0.001838
14	Geo_BBoxWidth	0.000327
15	Geo_AspectRatio	-0.001082
16	B01 (Coastal Aerosol)	-0.001518
17	B06 (Red Edge 2)	-0.001565
18	Geo_BBoxHeight	-0.002113
19	NDSI	-0.006397

Positive importance indicates a drop in F1-score upon permutation. Negative values suggest permutation improved performance, implying the feature might be noisy or redundant for this specific non-augmented model configuration.

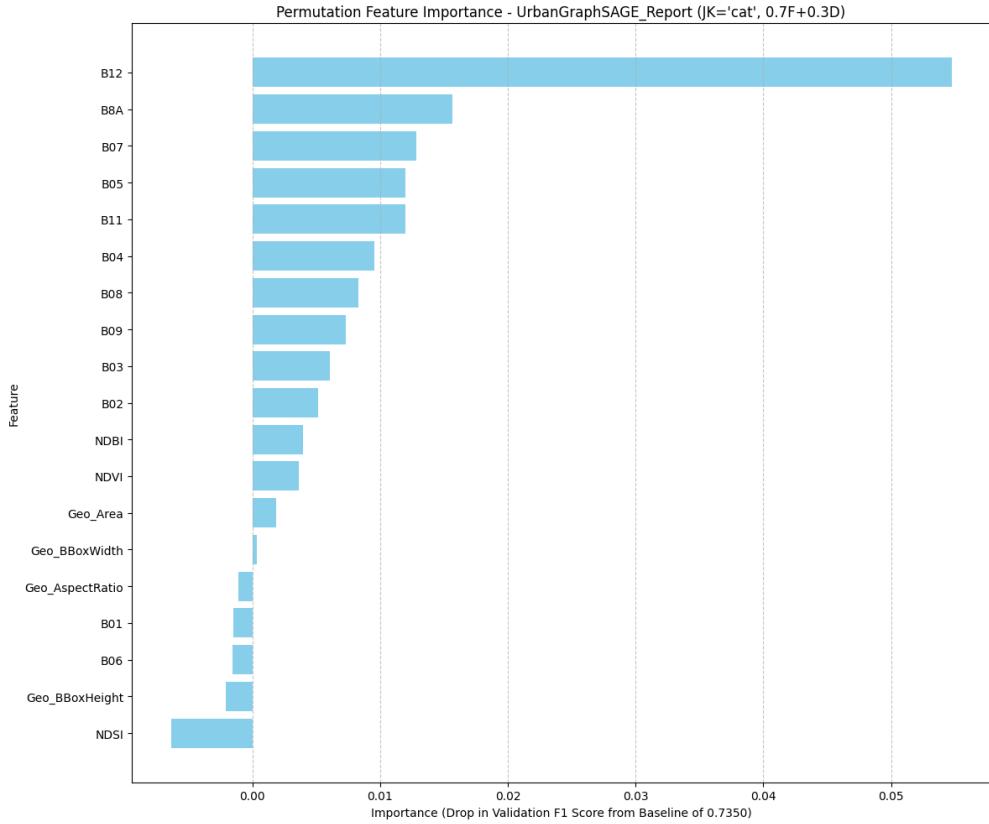


Figure 4.6: Permutation feature importance for the non-augmented UrbanGraphSAGE model ($JK='cat'$, $0.7F+0.3D$). Importance is measured as the drop in Validation F1-Score from a baseline of 0.7350 when each feature is permuted. Features are ranked from most to least important (top to bottom). (Source: Author, generated from permutation importance analysis detailed in Section 3.4.3)

The results highlighted that spectral bands, particularly from the Short-Wave Infrared (SWIR) and Red Edge regions, were the most influential features for this non-augmented model configuration. Specifically, B12 (SWIR 2), B8A (Narrow NIR), B07 (Red Edge 3), B05 (Red Edge 1), and B11 (SWIR 1) constituted the top 5 most important features. These bands are known for their effectiveness in distinguishing various land cover types, including built-up areas, due to their sensitivity to material properties and vegetation characteristics. Commonly used spectral indices such as NDBI and NDVI, while showing positive importance, ranked lower than many individual raw spectral bands. Geometric features (Geo_Area, Geo_BBoxWidth, Geo_AspectRatio, Geo_BBoxHeight) generally exhibited very low positive or even negative importance scores (e.g., Geo_AspectRatio at -0.001082 , NDSI at -0.006397), suggesting that for this superpixel-based classification with the non-augmented model, the precise shape or size attributes of superpixels were less discriminative than their aggregated spectral content. Features with negative importance scores (where permuting them led to a slight increase in model performance) might indicate redundancy or that they introduced noise

in the context of this specific non-augmented model and dataset.

To further investigate the practical implications of these feature importances, the non-augmented UrbanGraphSAGE model was retrained using progressively reduced feature sets, based on the ranking obtained. The performance of these models on the test set is summarized in Table 4.7.

Table 4.7: Summary of Test Set Performance for Non-Augmented UrbanGraphSAGE Model with Reduced Feature Sets (Baseline Test F1-Score with all 19 Features: 0.7297).

No. of Features	Test Accuracy	Test Precision	Test Recall	Test F1-Score	F1 Change from Baseline
19 (Baseline)	0.8095	0.6203	0.8859	0.7297	N/A
Top 17	0.7898	0.5886	0.9154	0.7165	-0.0132
Top 15	0.7868	0.5846	0.9167	0.7139	-0.0157
Top 10	0.7746	0.5683	0.9282	0.7050	-0.0247
Top 5	0.7757	0.5707	0.9154	0.7031	-0.0266

As shown in Table 4.7, reducing the feature set generally led to a decrease in the Test F1-Score. Removing the two least important features (those with negative permutation importance, resulting in a 17-feature set) led to a modest F1-score drop of 0.0132. More aggressive reductions to the top 10 and top 5 features resulted in more substantial performance decreases (0.0247 and 0.0266 respectively). This indicates that while some features might have low individual permutation importance or even appear detrimental in isolation for the non-augmented model, they can still contribute to the overall predictive power, likely through complex, non-linear interactions learned by the GNN. The decision to use all 19 features for the final spectrally augmented UrbanGraphSAGE model (discussed in Section 3.4.3 of Chapter 3) was based on the premise that augmentation would enable the model to better leverage the complete feature set, mitigating issues of noise or redundancy observed in the simpler non-augmented context and ultimately achieving superior performance.

4.1.6 Qualitative Results: Visual Assessment of Building Footprint Extraction

Beyond the quantitative metrics presented, a qualitative visual assessment of the extracted building footprints provides crucial insights into the model’s real-world performance, its strengths in handling diverse urban typologies, and specific areas where challenges remain. This assessment was performed by overlaying the

building footprint predictions generated by the best-performing model—the UrbanGraphSAGE trained with spectral augmentation and assessed OpenStreetMap (OSM) data—onto the original Sentinel-2 imagery for selected, representative sub-regions within the Algiers Area of Interest (AOI). The visual examples aim to showcase both successful delineations and instances where the model encountered difficulties, thereby offering a balanced perspective on its capabilities.

Examples of Successful Building Footprint Delineation

In several representative areas across the Algiers AOI, the augmented UrbanGraphSAGE model demonstrated a strong capability to accurately delineate building footprints under varying conditions of urban density and layout.

Figure 4.7 illustrates the model’s performance in an area characterized by high-density urban fabric, where buildings are closely packed with complex arrangements. Despite the challenging proximity of structures, the model was largely successful in identifying individual building footprints and capturing their general outlines. This suggests that the superpixel-based approach, combined with the GNN’s ability to learn from contextual neighborhood information, aided in distinguishing adjoined or nearly adjoined structures in these dense settings.



Figure 4.7: Example of successful building footprint extraction by the augmented UrbanGraphSAGE model in an area with high-density urban fabric. Predicted footprints (yellow overlay) are displayed on the Sentinel-2 RGB composite. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])

Similarly, Figure 4.8 showcases the model's effectiveness in an area with more moderate density and well-separated structures. In these scenarios, the model accurately captured the extents of both larger building blocks and smaller, detached units, demonstrating its adaptability to different building scales and layouts. The clear definition of boundaries suggests that the spectral and geometric features of the superpixels were highly discriminative for robust classification in these less congested environments.



Figure 4.8: Example of successful building footprint extraction by the augmented UrbanGraphSAGE model in an area with moderate-density urban fabric and well-separated structures. Predicted footprints (yellow overlay) are displayed on the Sentinel-2 RGB composite. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])

Examples of Common Challenges and Misclassifications

Despite its overall strong performance, the visual assessment also revealed specific scenarios where the model encountered challenges, leading to misclassifications. Understanding these error patterns is crucial for identifying the model's current limitations.

Figure 4.9 illustrates difficulties in accurately delineating very small or highly fragmented structures, a challenge primarily stemming from the 10-meter spatial resolution of the Sentinel-2 imagery. In such cases, individual small buildings might be represented by only a few superpixels or merged with surrounding non-building features, leading to omission errors where the model fails to detect them.



Figure 4.9: Example illustrating challenges in delineating very small or fragmented building footprints. The yellow overlay shows where the model has potentially missed small structures or struggled with their precise boundaries. (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])

Another observed challenge, depicted in Figure 4.10, was spectral confusion between building materials and certain non-building surfaces. The figure shows a region containing dark bare soil that the model has incorrectly classified as 'building' (a commission error). This type of misclassification occurs when non-building surfaces exhibit spectral signatures similar to those of some roofing materials (e.g., asphalt or dark tiles), causing the model to misinterpret them, even with contextual information from the GNN.



Figure 4.10: Example illustrating a commission error due to spectral confusion. The model has incorrectly classified a large area of dark bare soil as 'building' (yellow overlay). (Source: Author, Sentinel-2 imagery ©Copernicus Sentinel data [2024/2025])

This qualitative visual assessment complements the quantitative metrics by providing actionable intelligence on the model's real-world utility. While the results indicate strong performance, these visual examples highlight that challenges related to image resolution and spectral ambiguity persist. This detailed visual inspection is invaluable for identifying systematic error patterns and informs targeted recommendations for future work, such as the potential integration of higher-resolution data or more advanced feature engineering to tackle spectral confusion.

4.2 Analysis and Interpretation of Results

This section provides an in-depth analysis and interpretation of the quantitative and qualitative results presented in Section 4.1. The aim is to explore the underlying causes for the observed performance of the UrbanGraphSAGE model and its comparators, discuss the interplay between different methodological components, and understand the model's behavior in the context of the Algiers urban environment.

4.2.1 Performance Analysis of the Augmented UrbanGraph-SAGE Model

The augmented UrbanGraphSAGE model achieved a commendable Test F1-Score of 0.7579, demonstrating its overall efficacy for urban building footprint identification from Sentinel-2 imagery in Algiers. The model’s exceptional recall of 0.9192, albeit with a more moderate precision of 0.6448, warrants further discussion.

The high recall signifies the model’s strong ability to identify the vast majority of actual building superpixels, minimizing omission errors (only 63 False Negatives). This characteristic is largely attributable to the design of our custom loss function, `CombinedFocalDiceLoss`. The Dice Loss component directly optimizes for Intersection over Union (IoU), inherently pushing the model to maximize overlap with ground truth building areas, which often translates to higher recall. Simultaneously, the Focal Loss component, with an `alpha` parameter of approximately 0.7100, gave increased weight to the under-represented ‘building’ class, further encouraging the model to correctly identify these minority instances. This deliberate prioritization of recall is often desirable in applications like disaster damage assessment or urban expansion monitoring, where failing to detect an existing structure can have more severe consequences than some level of over-detection.

The moderate precision, reflected by 395 False Positives, indicates a tendency for the model to classify some non-building superpixels as buildings. These commission errors primarily arise from the inherent spectral complexity of urban environments. Surfaces such as bright, dry bare soil, certain types of roads, courtyards, or even some low, sparse vegetation in arid contexts can exhibit spectral signatures similar to those of common roofing materials (e.g., concrete, weathered metal) in Sentinel-2’s medium-resolution bands. While the GNN architecture, particularly the use of SAGEConv layers for aggregating neighborhood information and the JumpingKnowledge layer for multi-scale feature fusion, is designed to leverage spatial context to disambiguate such cases, these mechanisms cannot entirely overcome strong spectral similarities. The `mean` aggregator in SAGEConv, while robust to outliers, might smooth out subtle distinguishing features in highly heterogeneous superpixels. The JumpingKnowledge layer effectively combines local and global context, allowing the model to consider broader patterns beyond individual superpixel spectra, which likely contributed to its overall strong performance despite these challenges. However, if a non-building area is spectrally very similar to buildings and is surrounded by other spectrally ambiguous features, even contextual information may not be sufficient for perfect discrimination at this spatial resolution.

4.2.2 Impact of Assessed OSM Data and Spectral Augmentation on Model Robustness

Two critical methodological choices significantly contributed to the robustness and performance of the final UrbanGraphSAGE model: the rigorous assessment of OpenStreetMap (OSM) data for ground truth generation and the application of image-level spectral augmentation.

The multi-source data assessment process (cross-validating OSM with GOB and Overture Maps, followed by temporal NDVI-based stability checks for score 1 buildings) was instrumental in creating a higher-fidelity training dataset. By prioritizing buildings with multi-source agreement (scores 2 and 3) and by promoting temporally stable, non-vegetated score 1 buildings (predominantly from GOB in this study, indicating areas where OSM might have been incomplete), the training process was less affected by potential errors or omissions present in any single raw open dataset. This careful curation of ground truth labels (resulting in 88,638 reliable footprints) meant the GNN learned to associate image features with more accurate and representative building instances. This minimized the propagation of label noise into the model’s learned weights, leading to a more precise and generalizable classifier than would likely have been achieved using unassessed OSM data alone.

Spectral augmentation further enhanced model robustness and generalization, as evidenced by the F1-Score improvement of approximately 0.0282 (from a non-augmented Test F1 of 0.7297 to 0.7579 for the augmented model). By exposing the model to variations in brightness, contrast, and blur, it learned features more invariant to common real-world radiometric and atmospheric variations. This is crucial because Sentinel-2 imagery, even after Level-2A processing, can exhibit subtle differences across acquisition dates due to atmospheric haze, sun angle, and sensor calibration drifts. The augmentation forced the model to focus on more fundamental spectral and contextual patterns indicative of buildings, rather than overfitting to the specific radiometric characteristics of the original training images. The re-computation of superpixels and features for each augmented image instance, while computationally intensive, ensured that the learned graph representations were always consistent with the (augmented) pixel data, further strengthening this learning process. The combined effect of cleaner labels and more diverse training data created a self-reinforcing cycle, enabling the model to learn more discriminative features effectively.

4.2.3 Comparative Analysis of GNN Architectures (UrbanGraphSAGE vs. GCN vs. GAT)

The comparative analysis revealed that UrbanGraphSAGE (Test F1: 0.7579) and UrbanGCN_Enhanced (Test F1: 0.7535) performed comparably well and significantly outperformed UrbanGAT (Test F1: 0.6486) on this specific building footprint extraction task in Algiers, using the same augmented training data.

The close performance between UrbanGraphSAGE and UrbanGCN_Enhanced, despite their architectural differences, suggests that for this superpixel-based graph representation derived from Sentinel-2 data, the fundamental mechanism of aggregating local neighborhood information is highly effective. UrbanGCN_Enhanced, with its simpler message passing (averaging features from all direct neighbors) and skip connections, proved nearly as capable as UrbanGraphSAGE. The slight edge of UrbanGraphSAGE might be attributed to its inductive learning of aggregation functions (even if a simple 'mean' aggregator was chosen, the framework supports learned aggregators) and the powerful multi-scale feature fusion provided by its JumpingKnowledge 'cat' layer, which allows the final classifier to consider features from different levels of neighborhood aggregation simultaneously. The higher recall of UrbanGraphSAGE suggests its feature representation might be slightly more sensitive to capturing all instances of buildings, while GCN's slightly higher precision suggests it might be marginally better at avoiding false positives, a trade-off potentially influenced by subtle ways these architectures interact with the loss function.

The significant underperformance of UrbanGAT is noteworthy. GATs employ attention mechanisms to assign different importance weights to neighbors during aggregation. While theoretically powerful for capturing complex relationships, attention mechanisms can be prone to overfitting or becoming sensitive to noise in highly heterogeneous and potentially noisy datasets like superpixel graphs derived from real-world urban scenes. The Algiers urban landscape, with its diverse materials, shadows, and mixed superpixels, may have presented a challenging environment for the GAT model to learn stable and generalizable attention weights. The higher number of false positives (543 for GAT vs. 395 for UrbanGraphSAGE and 320 for GCN) supports the hypothesis that the GAT model might have struggled to robustly distinguish non-building urban features, possibly due to over-attending to certain misleading spectral or textural cues within superpixel neighborhoods. This suggests that for this specific application and data resolution, the more robust, averaging-based aggregation of GraphSAGE and GCN might be inherently more suitable or less prone to instability than the more fine-grained attention mechanism of the GAT model as configured. The similar number of trainable parameters

between UrbanGraphSAGE (19,585) and UrbanGCN_Enhanced (19,585) versus the lower count for UrbanGAT (10,433) indicates that model size alone was not the determining factor, pointing more towards architectural suitability for the task.

4.2.4 Interpretation of Feature Importance and Selection (Non-Augmented Model Context)

The permutation feature importance study conducted on the non-augmented UrbanGraphSAGE model (baseline Validation F1: 0.7350, detailed in Section 4.1.5 and Table 4.6) provided valuable insights into feature discriminability, even though the final model utilized all 19 features with augmentation.

The dominance of Short-Wave Infrared (B12, B11) and Red Edge (B8A, B07, B05) bands as the most influential features underscores their critical role. SWIR bands are particularly effective at differentiating artificial built-up surfaces from bare soil and vegetation due to distinct reflectance properties related to material composition and moisture content. Red Edge bands, sensitive to chlorophyll content and vegetation health, are crucial for accurately separating vegetated areas from non-vegetated (potentially urban) areas, thereby indirectly but significantly aiding in building detection by reducing confusion with green spaces.

The relatively lower positive importance of standard spectral indices like NDVI and NDBI, compared to individual raw spectral bands, suggests that the GNN model might be capable of learning more complex, direct relationships from the raw multi-band spectral information than what is captured by these pre-computed, often two-band ratio-based indices. The GNN, through its layers of aggregation and transformation, can potentially derive more nuanced spectral representations.

The very low or negative importance of most geometric features (Geo_Area, Geo_BBoxWidth, Geo_AspectRatio, Geo_BBoxHeight) and the NDSI for the non-augmented model is particularly interesting. It implies that for superpixel-based classification with this simpler training regime, the aggregated spectral content of a superpixel was far more dominant in the classification decision than its precise shape, size, or soil-related characteristics (as captured by NDSI). The negative importance of NDSI and Geo_AspectRatio even suggested that their inclusion might have been detrimental to the non-augmented model's performance, possibly by introducing noise or irrelevant information that confused the learning process.

The fact that reducing the feature set from 19 to the top 5 (Test F1 drop from 0.7297 to 0.7014) resulted in a noticeable performance decrease, even for the non-augmented model, indicates that even features with lower individual permutation importance contribute to the model's overall predictive power, likely

through non-linear interactions. The decision to retain all 19 features for the final spectrally augmented UrbanGraphSAGE model, which significantly outperformed the non-augmented baseline (Test F1 0.7579 vs. 0.7297), is further justified by this. Spectral augmentation likely enabled the model to better leverage the entire feature space, allowing it to learn more robust patterns and potentially find utility even in features that appeared less significant or noisy in the more limited, non-augmented training context. The increased data diversity may help the model disentangle complex feature interactions and reduce sensitivity to the idiosyncrasies of any single feature.

4.2.5 Observed Challenges and Limitations in Building Footprint Extraction

The qualitative visual assessment (Section 4.1.6) and the quantitative metrics (particularly precision scores and False Positives) highlighted several persistent challenges and limitations in the building footprint extraction process within the Algiers study area, even with the best-performing augmented UrbanGraphSAGE model.

- **Delineation of Small and Irregularly Shaped Buildings:** The inherent 10-meter spatial resolution of the primary Sentinel-2 bands limits the ability to accurately delineate very small structures (e.g., small sheds, kiosks, or fragmented informal dwellings) or buildings with highly complex, irregular outlines. Such structures may be smaller than a single pixel or span only a few pixels, leading to mixed superpixels where the building signature is diluted by surrounding non-building surfaces. This often results in omission errors (FNs) or highly generalized footprint shapes if detected.
- **Spectral Confusion with Non-Building Impervious Surfaces:** A significant source of commission errors (FPs) was the spectral similarity between certain building materials (e.g., light-colored concrete, metallic roofs, some types of asphalt) and other non-building impervious surfaces like roads, large paved areas (courtyards, parking lots), and bright, dry bare soil, especially in peripheral or less vegetated zones. While the GNN's use of context helps, strong spectral overlap can still lead to misclassification.
- **Impact of Dense Urban Fabric and Adjacency:** In highly dense urban areas like parts of the Casbah or tightly packed residential blocks, accurately separating individual, adjoined building footprints is challenging. Superpixels in these regions may struggle to conform perfectly to individual building boundaries, sometimes merging adjacent structures or including shared

walls/spaces. This can lead to an underestimation of building counts or less precise boundary delineation, though the overall area might still be correctly identified as 'built-up'.

- **Shadows:** Despite preprocessing efforts to mask clouds and their shadows, shadows cast by tall buildings or topographic features remain a challenge. Shadows significantly alter the spectral reflectance of underlying surfaces, making shadowed portions of buildings appear spectrally different from their sunlit parts, and potentially causing them to be misclassified as non-building or omitted. Shadowed non-building areas can also sometimes be confused with certain building materials.
- **Superpixel Segmentation Imperfections:** While SLIC generally adheres well to object boundaries, the segmentation is not perfect. Superpixels can sometimes cross actual building boundaries or average heterogeneous areas, especially in complex transition zones. The features extracted from such imperfect superpixels will inherently contain mixed signals, affecting the GNN's classification accuracy at these nodes.

These observed challenges are a combination of limitations stemming from the input data resolution (Sentinel-2), the inherent complexities of the urban environment in Algiers, and the current capabilities of the superpixel segmentation and GNN modeling pipeline. While the proposed methodology demonstrates strong performance, these aspects represent areas for potential future refinement and investigation.

4.3 Discussion (Contextualization and Broader Implications)

The results presented and analyzed in the preceding sections provide valuable insights into the application of Graph Neural Networks, specifically the Urban-GraphSAGE architecture, for urban building footprint identification using Sentinel-2 imagery and assessed OpenStreetMap data in Algiers. This section discusses these findings in a broader context, directly addressing the initial research questions, highlighting the contributions of this study to the field, exploring practical implications for urban planning and management in Algiers, acknowledging the inherent limitations of the research, and proposing promising avenues for future work.

4.3.1 Answering Research Questions

The methodology and subsequent results effectively address the key research questions posed at the outset of this study (Section 3.1):

RQ1: GNN Efficacy for Footprint Extraction: The UrbanGraphSAGE model, leveraging superpixel-derived features and neighborhood relationships from Sentinel-2 imagery, demonstrated considerable efficacy in identifying urban building footprints in Algiers, achieving a robust Test F1-Score of 0.7579. This indicates that GNNs, particularly architectures like GraphSAGE that can learn inductive representations from local neighborhoods and aggregate multi-scale information via mechanisms like JumpingKnowledge, are indeed a viable and effective approach for this complex semantic segmentation task even with medium-resolution satellite data. Its high recall (0.9192) is particularly noteworthy for applications requiring comprehensive building detection.

RQ2: Impact of Ground Truth Quality: The rigorous multi-source assessment of OSM data (cross-validation with GOB and Overture Maps) combined with temporal stability analysis using NDVI proved crucial. While a direct quantitative comparison against a model trained on raw, unassessed OSM data was not the primary focus of the final reported experiments (due to the known variability of raw OSM), the detailed data assessment pipeline itself (Section 3.2.3 and 3.2.3) demonstrated the substantial filtering and refinement achieved. For instance, promoting 3,921 Score 1 buildings to Medium_Score1_StableNonVeg status while excluding others like Stable_Vegetated (9 buildings) from the final 88,638 reliable footprints indicates a significant curation process. This curated, higher-confidence ground truth dataset is foundational to the trustworthy training and evaluation of the GNN, likely contributing significantly to the achieved performance by reducing label noise and ensuring the model learns from more accurate building representations.

RQ3: Contribution of Spectral Augmentation: Image-level spectral data augmentation demonstrably improved the performance and generalization of the UrbanGraphSAGE model. The F1-Score increased by approximately 0.0282 (from a non-augmented Test F1 of 0.7297 to 0.7579 for the augmented model). This quantitative improvement validates spectral augmentation as an effective regularization technique, forcing the model to learn features more invariant to realistic variations in illumination, atmospheric conditions, and sensor noise, thereby enhancing its robustness for real-world applications.

RQ4: Comparative GNN Performance: When compared under identical conditions (assessed ground truth, spectral augmentation), UrbanGraphSAGE

(Test F1: 0.7579) performed marginally better than UrbanGCN_Enhanced (Test F1: 0.7535) and significantly outperformed UrbanGAT (Test F1: 0.6486). This suggests that for this task, while robust local feature aggregation (common to both GraphSAGE and GCN) is highly effective, the inductive learning and multi-scale feature fusion of GraphSAGE may offer a slight advantage. The underperformance of GAT suggests that its attention mechanism might be less suited to the inherent noise and heterogeneity of superpixel graphs derived from Sentinel-2 imagery in complex urban scenes, or may require more extensive tuning.

RQ5: Feature Salience in Algiers (Non-Augmented Context): The permutation feature importance study on the non-augmented UrbanGraphSAGE model identified Short-Wave Infrared (SWIR: B12, B11) and Red Edge (B8A, B07, B05) bands as the most discriminative features. This underscores their critical role in distinguishing built-up areas from other land covers in the Algiers context, likely due to their sensitivity to material properties and vegetation stress. While spectral indices and geometric features showed lower individual importance for that model, the decision to use all 19 features in the final augmented model was justified by its superior overall performance, suggesting augmentation enables better utilization of the complete feature set.

4.3.2 Contribution to the Field

This research contributes to the growing body of knowledge in remote sensing and geospatial artificial intelligence for urban monitoring in several ways:

- **Novel Methodological Framework:** It presents and validates a synergistic framework combining advanced GNNs (UrbanGraphSAGE) with superpixel segmentation of Sentinel-2 imagery, a rigorous multi-source assessment pipeline for OSM-derived ground truth (incorporating GOB and Overture data), and effective spectral data augmentation. This holistic approach addresses key challenges in data quality, model robustness, and contextual feature learning for building footprint extraction.
- **Empirical Evidence for GNNs in Urban Mapping:** The study provides further empirical evidence for the effectiveness of GNNs, particularly GraphSAGE, in a complex, real-world urban environment using medium-resolution satellite data. The detailed comparison with GCN and GAT offers insights into architectural suitability for superpixel-based classification tasks.
- **Robust Ground Truth Generation Strategy:** The proposed multi-source cross-validation and temporal consistency checking pipeline for OSM

data offers a replicable and robust strategy for generating higher-quality ground truth labels from openly available VGI and global building datasets. This is a significant contribution, as ground truth quality is a persistent bottleneck in applying supervised learning to remote sensing tasks.

- **Demonstrated Efficacy of Spectral Augmentation:** The quantifiable improvement due to spectral augmentation reinforces its importance as a data-efficient regularization technique, particularly valuable when diverse labeled datasets are scarce.
- **Insights for North African Urban Contexts:** The feature importance analysis and qualitative results provide specific insights into the discriminative power of Sentinel-2 spectral bands for building identification in a North African urban context like Algiers, highlighting the utility of SWIR and Red Edge bands.

4.3.3 Implications for Urban Planning and Management in Algiers

The accurate and relatively up-to-date building footprint maps generated by the UrbanGraphSAGE model, leveraging freely available Sentinel-2 imagery and open building datasets, hold substantial practical implications for urban planning and management in a dynamic city like Algiers:

- **Urban Growth Monitoring and Sprawl Analysis:** The outputs can serve as a reliable baseline for tracking urban expansion over time, identifying new construction areas, monitoring changes in building density, and analyzing patterns of urban sprawl. This is crucial for informed land-use planning and managing the city's rapid urbanization.
- **Informal Settlement Mapping and Upgrading:** The model's high recall is particularly valuable for identifying and delineating informal settlements, which often lack official records. This information is essential for governmental and non-governmental organizations involved in targeted service provision, infrastructure planning, and improving living conditions in these vulnerable areas.
- **Infrastructure and Public Service Planning:** Precise knowledge of building locations and extents is critical for optimizing the planning and delivery of urban infrastructure (roads, water, electricity, sanitation) and public services (schools, healthcare facilities, emergency services).

- **Disaster Risk Reduction and Management:** Accurate building inventories are fundamental for various stages of disaster management, including pre-disaster vulnerability assessment (e.g., identifying buildings in flood-prone or geologically unstable areas), and post-disaster rapid damage assessment to guide response and recovery efforts.
- **Enhanced Population Estimation:** When combined with appropriate demographic data or dasymetric mapping techniques, building footprints can contribute to more accurate and spatially disaggregated population estimates, supporting better resource allocation and evidence-based policy formulation.

The methodology's reliance on open data sources also makes it potentially replicable and scalable, offering a cost-effective approach for continuous urban monitoring in Algiers and similar urban contexts.

4.3.4 Limitations of the Study

Despite the promising results, this study acknowledges several limitations inherent in the methodology, data, and scope of the research:

- **Ground Truth Reliance and Residual Uncertainty:** While the multi-source assessment significantly improved ground truth quality, the primary reliance on OSM, GOB, and Overture data means that any shared systematic biases or undiscovered errors in these large-scale datasets could still propagate. Even rigorously assessed open-source data may not perfectly capture all ground conditions or rapid temporal changes.
- **Spatial Resolution of Sentinel-2:** The 10-meter effective resolution for key bands, while good for large-area mapping, inherently limits the ability to accurately delineate very small, narrow, or highly complex building structures, leading to mixed superpixels and potential misclassifications as observed in the qualitative assessment.
- **Specificity to Algiers AOI:** The model was trained and optimized for the specific urban morphology, building typologies, and environmental conditions of the Algiers study area. Direct transferability to other urban areas with significantly different characteristics (e.g., different building materials, vegetation types, atmospheric conditions) may require further fine-tuning or adaptation.

- **Superpixel Segmentation Imperfections:** While SLIC is effective, no superpixel algorithm is perfect. Occasional instances of over- or under-segmentation, or boundaries not perfectly aligning with true object edges, can impact the purity of node features and labels, subsequently affecting GNN performance.
- **Scope of GNN Architectures and Hyperparameter Tuning:** Only a few GNN architectures (UrbanGraphSAGE, GCN, GAT) were explored. Other advanced GNN variants might offer different performance characteristics. Furthermore, while key hyperparameters were set based on previous work and experimentation (e.g., SLIC parameters), an exhaustive hyperparameter optimization for all GNN models was beyond the computational scope of this Master’s thesis.
- **Feature Importance Context:** The detailed permutation feature importance study was conducted on a non-augmented version of the UrbanGraphSAGE model. While insightful, the relative importance of features might shift in the context of a model trained with extensive spectral augmentation.
- **Shadow and Dense Urban Fabric Challenges:** As highlighted in the qualitative analysis, accurately classifying buildings in areas with significant shadow contamination or extremely dense, contiguous urban fabric remains a challenge even for the GNN approach.

4.3.5 Recommendations for Future Work

Based on the findings and limitations of this study, several avenues for future research are recommended to further advance the field of automated urban building footprint identification using GNNs:

- **Exploring Advanced GNN Architectures and Hyperparameters:** Investigate more sophisticated GNN architectures, such as those incorporating improved attention mechanisms with better regularization, deeper GCN/GraphSAGE variants with residual connections, or hybrid models. Conduct more extensive hyperparameter tuning, potentially using Bayesian optimization or similar advanced techniques.
- **Integration of Multi-Modal Data Sources:** Enhance feature representation by integrating complementary remote sensing data. For instance, LiDAR data could provide precise height information to differentiate buildings from flat impervious surfaces. Synthetic Aperture Radar (SAR) data, with

its all-weather capability and sensitivity to structure, could also be valuable, especially in persistently cloudy regions. Exploring fusion techniques for higher-resolution optical imagery (e.g., from PlanetScope or commercial providers) for specific areas of interest could address issues with small building delineation.

- **Advanced Data Augmentation Techniques:** Beyond image-level spectral augmentation, explore graph-level augmentation techniques such as node feature masking, edge perturbation (EdgeDrop), or sub-graph sampling to further improve model robustness and generalization. Generative Adversarial Networks (GANs) could also be investigated for creating synthetic but realistic training samples.
- **Transferability and Domain Adaptation Studies:** Conduct studies to assess the transferability of the trained models to other diverse urban environments (different cities, climatic zones, development patterns) and evaluate their performance with data from different but similar sensors (e.g., Landsat 8/9). Develop domain adaptation techniques to improve model performance when applied to new, unseen target domains.
- **Enhanced OSM Quality Assessment and Ground Truth Generation:** Develop more automated and refined pipelines for OSM quality assessment and ground truth generation, potentially incorporating machine learning for discrepancy resolution or more sophisticated temporal change detection algorithms to keep ground truth current.
- **Re-running Permutation Feature Importance on the Fully Augmented Model:** Performing a comprehensive feature importance study on the final, spectrally augmented UrbanGraphSAGE model would provide more definitive insights into feature contributions in the presence of rich data augmentation and could guide more informed feature engineering or selection for future iterations.
- **Improved Handling of Shadows and Complex Geometries:** Investigate specific modules or GNN modifications to better handle areas affected by shadows or to improve the delineation of highly complex or very small building geometries, possibly through multi-task learning or by incorporating explicit shape priors.

Chapter 5

Conclusion

This chapter provides a comprehensive summary of the research undertaken in this thesis, reiterating the core objectives, synthesizing the key findings, outlining the significant contributions, acknowledging the study's limitations, and offering concluding remarks on the broader implications of the work.

5.1 Restatement of Research Objectives and Scope

This thesis addressed the persistent challenge of accurate and efficient urban building footprint identification, a critical task for sustainable urban planning and development, particularly in rapidly urbanizing regions such as Algiers. The rapid and often uncontrolled expansion of the built environment in cities like Algiers, a rapidly growing capital city, necessitates robust quantitative tools for effective urban planning and climate action. Traditional manual or semi-automated methods for building footprint extraction are often labor-intensive, time-consuming, and struggle to keep pace with dynamic urban expansion, highlighting a significant gap that advanced automated solutions, as explored herein, can fill. The research undertaken aimed to overcome these limitations by leveraging advanced machine learning techniques and readily available open geospatial data. The practical relevance of this work extends beyond academic inquiry, directly linking to real-world urban planning challenges and the need for evidence-based decision-making in a city facing dual challenges of urbanization and climate change.

The primary objectives of this thesis were threefold:

- To develop and rigorously evaluate a novel Graph Neural Network (GNN) framework, termed UrbanGraphSAGE, specifically designed for urban building footprint identification from medium-resolution Sentinel-2 satellite imagery.
- To systematically assess the impact of quality-controlled OpenStreetMap

(OSM) data as a robust ground truth source and to quantify the benefits of image-level spectral data augmentation on the performance of the UrbanGraphSAGE model.

- To comparatively analyze the performance of UrbanGraphSAGE against other prominent GNN architectures, such as Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), within the context of urban mapping in Algiers.

The scope of this study was focused on the metropolitan area of Algiers, Algeria, utilizing Sentinel-2 multispectral imagery as the primary remote sensing data source. The methodological approach centered on superpixel-based GNNs, which convert image segments into graph structures to capture intricate spatial relationships and contextual information, thereby moving beyond pixel-level analysis.

5.2 Summary of Key Findings

The research conducted successfully developed and evaluated the UrbanGraphSAGE framework, demonstrating its efficacy in identifying urban building footprints within the Algiers study area. The key findings, derived from extensive experimentation and rigorous analysis as presented in Chapter 4, are summarized below:

- **Performance of the Augmented UrbanGraphSAGE Model:** The primary UrbanGraphSAGE model, trained with assessed OpenStreetMap (OSM) data and enhanced by image-level spectral data augmentation, exhibited strong performance. Specifically, on the test set, the model achieved:
 - An F1-Score of **0.7579**.
 - A Recall of **0.9192**.
 - A Precision of **0.6448**.
 - An Accuracy of **0.8296**.

These metrics, detailed in Section 4.1.1, indicate a robust capability in minimizing the omission of actual building footprints, a critical factor for comprehensive urban mapping, while maintaining a reasonable level of precision. This performance underscores the effectiveness of the synergistic combination of the GraphSAGE architecture, superpixel-based image representation, and strategic data handling.

- **Positive Impact of Assessed OpenStreetMap Data:** The meticulous quality control and multi-source cross-validation (with Google Open Buildings and Overture Maps) of OSM data, followed by temporal stability checks (detailed in Section 3.2.3 and 3.2.3), proved to be a critical factor in generating reliable ground truth. The data assessment process demonstrably filtered and refined potential building footprints, for example, by promoting 3,921 temporally stable, non-vegetated score 1 buildings to a higher trustworthiness category while excluding others (e.g., 9 identified as `Stable_Vegetated`) from the final 88,638 reliable footprints used for ground truth generation. This careful curation is understood to have qualitatively improved the model’s learning process by reducing label noise, thereby contributing to more accurate and spatially coherent building footprint detections. This highlights that rigorous data preparation is not merely a preliminary step but a fundamental determinant of model learning efficacy and generalizability in remote sensing applications utilizing VGI.
- **Quantitative Benefits of Spectral Data Augmentation:** Image-level spectral data augmentation yielded a clear and quantifiable improvement in model performance, as discussed in Section 4.1.3. The UrbanGraphSAGE model trained with augmentation showed an increase in the Test F1-Score of **0.0282** compared to its non-augmented counterpart (which had a Test F1-Score of **0.7297**). This finding validates spectral augmentation as an effective technique for enhancing model robustness and generalization by exposing the model to a wider range of spectral conditions, thereby improving its ability to perform consistently on unseen data.
- **Comparative Performance Against Other GNN Architectures:** When compared under identical training conditions (using assessed OSM data and spectral augmentation, detailed in Section 4.1.2), UrbanGraphSAGE (Test F1: **0.7579**) performed marginally better than UrbanGCN_Enhanced (Test F1: **0.7535**) and significantly outperformed UrbanGAT (Test F1: **0.6486**). This suggests that while robust local feature aggregation is key, the inductive learning and multi-scale feature fusion capabilities of UrbanGraphSAGE offer some advantages for this task. The results also indicate that complex attention mechanisms, as in GAT, may not always translate to superior performance in noisy, heterogeneous urban environments without extensive tuning or specific architectural adaptations for this specific dataset and task.
- **Insights from SLIC Parameterization and Feature Importance (Non-Augmented Model Context):** The SLIC parameter experimentation

(summarized in Section 4.1.4) confirmed the importance of selecting appropriate values for `n_segments`, `compactness`, and `sigma` to generate meaningful superpixels. The chosen parameters (`n_segments=20000`, `compactness=1`, `sigma=1`), resulting in 17,916 actual superpixels for the original image, were found to provide a suitable trade-off for the Algiers imagery. The permutation feature importance study conducted on the non-augmented UrbanGraphSAGE model (Validation F1 baseline 0.7350, detailed in Section 4.1.5) highlighted the critical role of Short-Wave Infrared (SWIR) and Red Edge spectral bands, specifically **B12 (SWIR 2)**, **B8A (Narrow NIR)**, and **B07 (Red Edge 3)**, for distinguishing urban features in that context. While raw spectral bands generally outperformed pre-computed indices and geometric features in individual importance for that non-augmented model configuration, the overall strong performance of the final augmented model using all 19 features suggests that augmentation enables a more holistic utilization of the entire feature set.

5.3 Significance and Contributions of the Research

This research makes several significant contributions to the interconnected fields of remote sensing, geospatial artificial intelligence (GeoAI), and urban studies, particularly concerning the challenge of accurate and efficient urban monitoring in dynamic environments. The key contributions arising from the development and evaluation of the UrbanGraphSAGE framework for building footprint identification in Algiers are as follows:

- **Validation of a Novel Synergistic Framework for Urban Mapping:** A primary contribution of this thesis is the successful development, implementation, and validation of a synergistic framework that intelligently integrates multiple advanced techniques: Graph Neural Networks (specifically UrbanGraphSAGE), superpixel-based image segmentation of Sentinel-2 data, a rigorous multi-source assessment strategy for OpenStreetMap-derived ground truth (incorporating Google Open Buildings and Overture Maps data with temporal NDVI analysis), and image-level spectral data augmentation. This holistic approach demonstrates a pathway to overcome limitations inherent in individual components, leading to a more accurate and robust solution for urban building footprint identification from medium-resolution satellite imagery than might be achieved by these techniques in isolation. The framework itself offers a transferable blueprint for tackling complex geospatial feature extraction tasks.

- **Demonstration of a Robust and Replicable Ground Truth Generation Strategy:** The research empirically demonstrates a robust and effective strategy for generating higher-quality ground truth data from openly available VGI sources like OpenStreetMap. By implementing a meticulous multi-source cross-validation and temporal consistency checking pipeline, this study showed that the inherent quality variations in crowd-sourced data can be significantly mitigated, transforming OSM into a more reliable and scalable source for training advanced deep learning models. This addresses a critical bottleneck in many remote sensing applications where high-quality labeled datasets are scarce or expensive to acquire, thereby contributing to the democratization of high-precision urban mapping capabilities.
- **Empirical Evidence on GNN Performance for Urban Mapping with Medium-Resolution Satellite Data:** This study provides compelling empirical evidence supporting the strong performance of GNNs, and UrbanGraphSAGE in particular, for the complex task of urban building footprint identification using widely accessible medium-resolution Sentinel-2 data. While GNNs have shown promise in various domains, their application and detailed comparative evaluation for fine-grained urban feature extraction with this specific type of satellite imagery were less explored. The findings underscore the GNNs' inherent ability to effectively model spatial relationships and contextual information within superpixel graphs, which is crucial for accurately delineating irregular building shapes and navigating complex urban patterns. The comparative analysis against GCN and GAT architectures further offers valuable empirical insights into architectural suitability for such superpixel-based urban classification tasks.
- **Quantified Efficacy of Image-Level Spectral Data Augmentation:** The research provides clear quantitative evidence (an F1-Score improvement of 0.0282) for the benefits of applying image-level spectral data augmentation in the context of GNN-based building footprint extraction. This reinforces the importance of augmentation as a data-efficient regularization technique that enhances model robustness and generalization by exposing the model to a wider range of plausible spectral conditions, a particularly valuable asset when working with limited or inhomogeneously acquired training datasets.
- **Specific Insights for Urban Monitoring in North African Contexts:** Through the feature importance analysis (conducted on the non-augmented model) and the overall model performance evaluation in Algiers, this study provides specific insights into the discriminative power of different Sentinel-2

spectral bands for building identification in a North African urban setting. The highlighted importance of SWIR and Red Edge bands offers practical guidance for feature selection and sensor consideration in similar geographic and urban contexts. The successful application of the framework in Algiers also demonstrates its potential utility for addressing pressing urban monitoring needs in other rapidly urbanizing cities in the region.

Collectively, these contributions advance the understanding and application of advanced GeoAI techniques for tackling real-world urban challenges, paving the way for more automated, accurate, and scalable urban monitoring systems.

5.4 Limitations of the Study

Despite the significant contributions and promising results presented in this thesis, it is important to acknowledge several inherent limitations that define the boundaries of the current research and offer perspectives for future investigation. These limitations, discussed in more detail in Section 4.3.4, are concisely summarized as follows:

- **Reliance on Open Data and Inherent Resolution Constraints:** The methodology's foundation on open-source Sentinel-2 imagery and OpenStreetMap (OSM) data, while promoting accessibility and reproducibility, also introduces limitations. The 10-meter effective spatial resolution of key Sentinel-2 bands restricts the ability to precisely delineate very small or intricately shaped building footprints. Similarly, despite rigorous multi-source assessment, the VGI nature of OSM data means that residual uncertainties regarding completeness, positional accuracy, and temporal currency can persist and subtly influence model training and evaluation.
- **Geographic Specificity of the Study Area:** The research was conducted within a specific Area of Interest (AOI) in Algiers, Algeria. While chosen for its diverse urban characteristics, the findings and the trained model's optimal performance may not be directly transferable to other global urban environments with significantly different architectural styles, building materials, vegetation typologies, or atmospheric conditions without further adaptation or fine-tuning.
- **Scope of Graph Neural Network Exploration:** This study focused on the UrbanGraphSAGE architecture and its comparison with standard GCN and GAT models. The rapidly evolving field of GNNs offers a wider array

of architectures (e.g., with more complex attention mechanisms, dynamic graph capabilities, or deeper structures) that were not explored within the scope of this work. Furthermore, while key hyperparameters were carefully selected, an exhaustive optimization across the entire parameter space for all evaluated models was not feasible.

- **Superpixel Segmentation Artifacts:** The superpixel segmentation process, while crucial for the graph-based approach, can occasionally introduce artifacts where segment boundaries do not perfectly align with true object edges, particularly in highly complex or transitional urban zones. Such imperfections can influence the purity of node features and labels, subsequently affecting the GNN’s classification precision at these boundaries.
- **Context of Feature Importance Analysis:** The detailed permutation feature importance analysis was conducted on a non-augmented version of the UrbanGraphSAGE model. While providing valuable insights, the relative importance and interaction of features might differ in the context of the final model trained with comprehensive spectral augmentation, a specific investigation which was beyond the immediate timeline of this research.

Acknowledging these limitations is crucial for a balanced interpretation of the research outcomes and for identifying targeted areas where future work can build upon the foundations established by this thesis.

5.5 Concluding Remarks

This thesis has successfully developed and rigorously evaluated a novel framework integrating Graph Neural Networks, superpixel segmentation, quality-assessed OpenStreetMap data, and spectral augmentation for the purpose of urban building footprint identification from Sentinel-2 satellite imagery. The research undertaken has demonstrated the significant potential of this synergistic approach to address the persistent challenges associated with accurate and efficient urban monitoring, particularly within the dynamic and morphologically complex environment of Algiers.

The UrbanGraphSAGE model, at the core of this framework, achieved commendable performance, underscoring the capability of GNNs to effectively learn from graph-structured representations of geospatial data. The meticulous data assessment strategy, particularly the multi-source cross-validation of OSM data, proved indispensable in generating reliable ground truth, highlighting the critical importance of data quality in supervised learning applications for remote sensing.

Furthermore, the quantifiable benefits derived from image-level spectral augmentation have underscored its value as a practical technique for enhancing model robustness and generalization, especially when dealing with the inherent variability of satellite imagery.

The findings of this study contribute not only to the theoretical understanding of GNN applications in urban remote sensing but also offer tangible methodological advancements with practical implications. The ability to automatically and precisely map building footprints at scale, utilizing freely available satellite imagery and open data sources (albeit carefully curated), empowers urban authorities and planners with critical data for informed decision-making. This is particularly relevant for addressing pressing urban challenges in rapidly developing cities like Algiers, from monitoring urban expansion and managing informal settlements to supporting infrastructure development and contributing to sustainable development initiatives.

While acknowledging the limitations inherent in the study's scope and data sources, this work underscores the transformative power of advanced geospatial Artificial Intelligence in enhancing our capacity for automated urban monitoring. The research frames the developed methodology not merely as a technical achievement but as a foundational step towards a future where intelligent systems play an increasingly vital role in fostering smart city development and evidence-based climate action. The methodologies and insights presented herein lay a strong groundwork for continued innovation, paving the way for more intelligent, responsive, and sustainable urban futures. The journey to perfect automated interpretation of complex urban landscapes is ongoing, but this thesis represents a meaningful stride forward in that endeavor.

Bibliography

- [1] G. F. Hughes, “The mean accuracy of statistical pattern recognizers,” *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.
- [2] J. W. Rouse, R. Haas, J. Schell, and D. Deering, “Monitoring vegetation systems in the great plains with erts,” *NASA special publication*, vol. 351, p. 309, 1974.
- [3] C. J. Tucker, “Red and photographic infrared linear combinations for monitoring vegetation,” *Remote sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.
- [4] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *Proceedings of the ninth IEEE international conference on computer vision*, IEEE, vol. 1, 2003, pp. 10–17.
- [5] Y. Zha, J. Gao, and S. Ni, “Use of normalized difference built-up index in automatically mapping urban areas from tm imagery,” *International journal of remote sensing*, vol. 24, no. 3, pp. 583–594, 2003.
- [6] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [7] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [8] M. Haklay and P. Weber, “Openstreetmap: User-generated street maps,” *IEEE Pervasive computing*, vol. 7, no. 4, pp. 12–18, 2008.
- [9] J. Bennett, “Names and naming in openstreetmap,” *The Cartographic Journal*, vol. 47, no. 2, pp. 130–140, 2010.
- [10] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [11] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

- [12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” in *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, IEEE, 2012, pp. 2274–2282.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems 25*, 2012.
- [14] W. Liao, L. Zhang, and H. Shen, “K-nearest neighbor based methods for remote sensing image classification,” *Science China Information Sciences*, vol. 55, pp. 2114–2124, 2012.
- [15] OpenStreetMap Foundation, *Open Data Commons Open Database License (ODbL)*, <https://opendatacommons.org/licenses/odbl/1-0/>, Accessed: June 10, 2025, 2012.
- [16] U. Surendran and V. Kumar, “K-means clustering in remote sensing,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 1, no. 10, pp. 772–775, 2012.
- [17] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [18] T. Blaschke, “Object-based image analysis for remote sensing,” *ISPRS journal of photogrammetry and remote sensing*, vol. 90, pp. 18–29, 2014.
- [19] T. Chen, M. Li, J. Li, and C. Wang, “Shadow-based building detection from single very high resolution remote sensing imagery,” *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083618, 2014.
- [20] H. Fan, A. Zipf, and Q. Fu, “Quality assessment for building footprints data on openstreetmap,” *International Journal of Geographical Information Science*, vol. 28, no. 12, pp. 2411–2431, 2014.
- [21] P. Neubert and P. Protzel, “Compact watershed and preemptive slic: On improving trade-offs of superpixel segmentation algorithms,” in *22nd International conference on pattern recognition*, IEEE, 2014, pp. 996–1001.
- [22] European Space Agency, “Sentinel-2 user handbook,” ESA, Tech. Rep., 2015, Accessed: June 10, 2025.
- [23] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

- [24] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [25] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Feature engineering for remote sensing image classification: A comprehensive review,” *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 4, pp. 24–48, 2015.
- [26] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [27] M. Belgiu and L. Drăguț, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.
- [28] B. B. Chaudhuri, B. U. Shankar, and B. V. S. A. Kumar, “Automatic building detection from satellite images using a machine learning approach,” in *Pattern recognition and machine intelligence*, Springer, 2016, pp. 239–245.
- [29] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [30] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, “Machine learning in geosciences and remote sensing,” *Geoscience Frontiers*, vol. 7, no. 1, pp. 3–10, 2016.
- [31] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 565–571.
- [32] P. Mooney and P. Corcoran, “Towards understanding the temporal accuracy of openstreetmap: A quantitative experiment,” in *Proceedings of the Academic Track, State of the Map 2016*, 2016.
- [33] D. Tuia, C. Persello, and L. Bruzzone, “Domain adaptation for the classification of remote sensing data: An overview of recent advances,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 42–57, 2016.
- [34] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

- [35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [36] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural message passing for quantum chemistry,” in *International conference on machine learning*, PMLR, 2017, pp. 1263–1272.
- [37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [38] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, “Sen2cor for sentinel-2,” in *Image and Signal Processing for Remote Sensing XXIII*, SPIE, vol. 10427, 2017, pp. 337–349.
- [39] H. Senaratne, A. Mobasher, A. Ali, C. Capineri, and M. Haklay, “A review of volunteered geographic information quality assessment methods,” *International Journal of Geographical Information Science*, vol. 31, no. 1, pp. 139–167, 2017.
- [40] X. X. Zhu et al., “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.
- [41] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, “Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network,” in *ISPRS congress*, 2018.
- [42] D. Stutz, A. Hermans, and B. Leibe, “Superpixels: An evaluation of the state-of-the-art,” *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.
- [43] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, “Representation learning on graphs with jumping knowledge networks,” in *International conference on machine learning*, PMLR, 2018, pp. 5453–5462.
- [44] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*, 2019.
- [45] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.
- [46] Google Research, *Google Open Buildings*, <https://sites.research.google/open-buildings/>, Accessed: June 10, 2025, 2021.

- [47] B. Herfort, S. Lautenbach, J. Porto de Albuquerque, J. Anderson, and A. Zipf, “Investigating the digital divide in openstreetmap: Spatio-temporal analysis of inequalities in global urban building completeness,” *International Journal of Geographical Information Science*, vol. 35, no. 10, pp. 1963–1990, 2021.
- [48] W. C. Jochem and A. J. Tatem, “Tools for mapping multi-scale settlement patterns of building footprints: An introduction to the r package foot,” *Journal of Open Source Software*, vol. 6, no. 60, p. 2998, 2021.
- [49] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 2021.
- [50] E. Schiefer, S. van der Linden, S. Schönbrodt-Stitt, and P. Hostert, “A review of remote sensing-based methods for urban and rural settlement mapping,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, p. 102509, 2021.
- [51] Y. Zheng, Y. Wang, and Q. Wang, “Deep learning-based semantic segmentation of urban features in vhr remote sensing imagery: A comprehensive review,” *Remote Sensing*, vol. 13, no. 4, p. 808, 2021.
- [52] C. Ayala, C. Aranda, and M. Galar, “Pushing the limits of sentinel-2 for building footprint extraction,” *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
- [53] H. Han, J. Kang, M. Li, and X. Li, “Superpixelgraph: Semi-automatic generation of building footprint through semantic-sensitive superpixel and neural graph networks,” *arXiv preprint arXiv:2304.05661*, 2023.
- [54] O. Kucuk, V. Eset, F. C. Akyon, E. Sertel, and G. Unal, “A novel spatiotemporal method for object-based land cover classification of satellite imagery using a graph neural network,” *Scientific Reports*, vol. 13, no. 1, p. 12798, 2023.
- [55] Overture Maps Foundation, *Overture Maps Data*, <https://overturemaps.org/>, Accessed: June 10, 2025, 2023.
- [56] D. Patel, P. Srivastava, A. Bhattacharya, and P. Pandey, “Improving built-up extraction using spectral indices and machine learning on sentinel-2 satellite data in mumbai suburban district, india,” *International Journal of Engineering and Technology Innovation*, vol. 13, no. 4, pp. 281–294, 2023.
- [57] Y. Shi, Q. Li, and X. Zhu, “Building footprint extraction with graph convolutional network,” *arXiv preprint arXiv:2305.04499*, 2023.

- [58] I. UNDP Global Centre for Technology and S. Development, *Deep Tech Series Vol. 10: Looking to the Skies for Land Development: How Space Tech Can Contribute to Sustainable City Development*, <https://www.undp.org/policy-centre/singapore/blog/deep-tech-series-vol-10-looking-skies-land-development-space-tech-sustainable-city-development>, Accessed: June 10, 2025, 2023.
- [59] S. Al Shaan and D. Hu, “Integrating machine learning and remote sensing in disaster management: A decadal review of post-disaster building damage assessment,” *Buildings*, vol. 14, no. 8, p. 2344, 2024.
- [60] H. de Arruda and S. Reia, “An openstreetmap derived building classification dataset for the united states,” *Scientific Data*, vol. 11, no. 1, pp. 1–12, 2024.
- [61] J. Liu, J. Wu, H. Xie, D. Xiao, and M. Ran, “Semantic segmentation of deep learning remote sensing images based on band combination principle: Application in urban planning and land use,” *Applied Sciences*, vol. 14, no. 17, p. 7499, 2024.
- [62] H. Wang, X. Zhang, and J. Li, “Fusion of aerial and satellite images for automatic extraction of building footprint information using deep neural networks,” *Applied Sciences*, vol. 16, no. 5, p. 380, 2024.
- [63] B. Herfort, S. Lautenbach, J. de Albuquerque, and A. Zipf, “Ai-generated buildings in openstreetmap: Frequency of use and differences from non-ai-generated buildings,” *International Journal of Digital Earth*, 2025.
- [64] W. Lunga et al., “Socially inclusive infrastructure for disaster risk reduction in urban planning: Insights from the sadc region,” *Frontiers in Built Environment*, vol. 11, p. 1586 040, 2025.
- [65] L. See, A.-M. Olteanu-Raimond, and C. C. Fonte, “Recent advances in volunteered geographic information (vgi) and citizen sensing,” *International Journal of Digital Earth*, vol. 18, no. 1, e2480220, 2025.

Appendix A

Profile of the Algerian Space Agency (ASAL)

A.1 Introduction

The Algerian Space Agency (ASAL) is a national public institution established under Presidential Decree No. 02-48 in 2002. ASAL is responsible for implementing the country's space policy and serves as the principal body overseeing Algeria's space activities. This profile provides a comprehensive study of ASAL's history, organizational structure, mission responsibilities, satellite infrastructure, applications in remote sensing, and international partnerships.

A.2 Presentation of the Host Organization: ASAL

ASAL plays a central role in executing Algeria's space strategy. Its mandate includes the promotion and sustainable development of space technologies to serve national development goals, including environmental monitoring, telecommunications, national security, and scientific advancement.

A.2.1 Historical Overview

ASAL's evolution is rooted in decades of scientific advancement:

- **1975:** Establishment of the National School of Geodetic Sciences, initiating formal education in space-related disciplines.
- **1987:** Founding of the National Center for Space Techniques (CNTS).
- **2002:** Creation of the Algerian Space Agency (ASAL).
- **2006:** Adoption of the first National Space Program (2006–2020).

- **2020:** Launch of the second National Space Program (2020–2040).

A.2.2 Organizational Structure

The Algerian Space Agency (ASAL) is structured around a central administration and several specialized operational entities, each with distinct missions aligned with the objectives of the National Space Program. These entities are strategically distributed across various regions of Algeria to optimize the development, operation, and application of space technologies.

- **Centre des Techniques Spatiales (CTS) – Arzew:** The CTS is dedicated to scientific and technical research in areas such as space technology (including sensors, radiometers, ground stations, and Earth observation instruments), aerospace remote sensing physics, satellite image processing methodologies, spatial geodesy, satellite navigation systems, radio astronomy, space altimetry, and geomatics.
- **Centre de Développement des Satellites (CDS) – Bir El Djir, Oran:** Inaugurated in 2012, the CDS is responsible for the design, development, integration, and testing of Algerian satellites. It comprises modern infrastructures, including workshops and laboratories, dedicated to satellite assembly and environmental testing. The center also focuses on quality assurance and collaborates with national industries in related technological fields.
- **Centre d'Exploitation des Systèmes de Télécommunications (CESTS) – Boughezoul and Bouchaoui:** The CESTS manages the exploitation and commercialization of telecommunications satellite services, notably those provided by Alcomsat-1. Its responsibilities include technical management of ground reception and control infrastructures, coordination with user sectors, and implementation of service commercialization policies.
- **Centre des Applications Spatiales (CAS) – Algiers:** The CAS implements actions related to the exploitation of satellites and systems resulting from space programs, in collaboration with various user sectors. It undertakes operational projects based on remote sensing and geographic information systems (GIS) in fields such as environment, natural risk management, agriculture, water resources, land use planning, urban development, geology, and Earth sciences.
- **Centre d'Exploitation des Systèmes de Télédétection Satellitaire (CESTS):** The CESTS serves as a ground station for receiving and pro-

cessing satellite imagery, supporting various applications in remote sensing and Earth observation.

- **Center for Telecommunication Systems Operation (COSTS):** The COSTS manages the operation and commercialization of telecommunications satellite services, notably those provided by Alcomsat-1.
- **École Nationale Supérieure des Sciences Géodésiques et des Techniques Spatiales (ENSGTS) – Arzew:** Affiliated with ASAL, the ENSGTS offers advanced academic and professional training in geodesy, spatial techniques, cartography, and GIS. The institution aims to produce highly qualified engineers to support the national space program.

A.2.3 Logo of the Algerian Space Agency (ASAL)



Figure A.1: ASAL Official Logo.

A.3 ASAL's Operational Domains and Strategic Engagements

A.3.1 ASAL Core Missions

ASAL's activities are structured around several key mission areas:

- **Developing a National Space Strategy:** Advising the government by proposing a clear and ambitious vision for the promotion and development of space activities, ensuring alignment with national needs and international commitments.

- **Establishing Space Infrastructure:** Creating and managing the resources—technical, human, and financial—necessary for maintaining an autonomous national capability in space.
- **Satellite Design, Deployment, and Operation:** Managing programs for the conception, launch, and operational control of satellites to serve various applications including environmental monitoring and telecommunications.
- **Promoting International Cooperation:** Building strategic partnerships through bilateral and multilateral agreements, engaging with global space agencies and research institutions to exchange expertise and collaborate on advanced projects.

A.3.2 Satellite Infrastructure and Programs

ASAL manages a fleet of satellites for telecommunications and Earth observation. A summary of key space technologies is provided in Table A.1.

Table A.1: ASAL Space Technologies Appropriation.

Satellite	Purpose	Specifications	Launch Date
ALSAT-1	Earth Observation	32 m Resolution, 3 Bands RGB	28/11/2002
ALSAT-2A	Earth Observation	2.5 m Resolution PAN, 10 m MS 4 Bands	12/07/2010
ALSAT-1B	Alsat-1 Twin	12 m Resolution, 24 m MS 4 Bands	26/09/2016
ALSAT-2B	Alsat-2A Twin	-	26/09/2016
ALSAT-1N	Technological demonstration and scientific research	Cube sat 3U	26/09/2016
ALCOMSAT-1	Telecommunications	24.8°W, 33 transponders	10/12/2017

A.3.3 Remote Sensing and Technological Applications

Remote sensing is at the heart of ASAL’s research and operational activities:

- **Advanced Image Processing:** ASAL’s research focuses on refining methods for capturing, processing, and analyzing satellite images to maximize the utility and quality of the data.

- **Environmental and Urban Monitoring:** High-quality satellite data supports disaster prevention, environmental management, and strategic urban development.
- **Precision Agriculture:** Real-time monitoring of crops and soil conditions enables efficient irrigation management and early detection of anomalies, thereby increasing agricultural productivity.

A.3.4 International Collaboration and Research Opportunities

ASAL's commitment to international cooperation creates numerous research and development opportunities:

- **Global Partnerships:** The agency actively forges collaborations with international space agencies and research centers, facilitating knowledge exchange and joint development projects.
- **Academic and Industrial Synergies:** Through training programs, internships, and collaborative research projects, ASAL enhances the skills of young researchers and fosters innovation within the national industrial base.
- **Support for Research Projects:** ASAL provides access to cutting-edge infrastructures and financial support, offering a stable environment for PhD candidates and researchers to explore advanced space technologies.

A.4 Conclusion

ASAL is the driving force behind Algeria's ambitions in space. With a robust infrastructure, a capable satellite fleet, and growing international partnerships, the agency continues to play a vital role in advancing Algeria's scientific, technological, and strategic capabilities in the space domain.