AI Integrated Web Application Development for OSM Change Detection: A Case Study of Luxembourg,Western Europe

Snigdha Dsouza [1], Gauri Deshpande [2], T.P. Singh [3]

[1] Post-graduate student, Symbiosis Institute of Geoinformatics (SIG), Symbiosis International (Deemed University) (SIU), Model Colony, Pune, Maharashtra, India - snigdhadsouza900@gmail.com
[2] Assistant Professor, Symbiosis Institute of Geoinformatics (SIG), Symbiosis International (Deemed University) (SIU), Model Colony, Pune, Maharashtra, India- gaurideshpande127@gmail.com
[3] Director, Symbiosis Institute of Geoinformatics (SIG), Symbiosis International (Deemed University) (SIU), Model Colony, Pune Maharashtra, India - tarunsingh@rediffmail.com

Keywords: Open Street Map (OSM), data-driven approach, Land Use Land Cover (LULC)

Abstract

The rapid pace of urban growth demands continuous observation, especially in compact, high-development areas like Luxembourg. To effectively analyse land use transformations, scalable tools and an open, data-driven approach are essential for maintaining accuracy and clear visualization. The traditional methods of using remotely sensed data are bulky and difficult to work with, which requires a more easy and dynamic aspect. The present research introduces an advanced approach by creating an SQL algorithm to show the change detection of land use categories for 2016 and 2024 considering built-up areas. An AI-powered web application designed to detect and interpret changes in built-up using Open Street Map (OSM) data. By incorporating modern technologies with AI, particularly Large Language Model (LLMs), this platform provides a simple interface for retrieval of statistical data, also querying and analysing the spatial trends. The accuracy of built-up changes derived from OSM was validated by comparing them with a supervised Land Use Land Cover (LULC) classification generated from Sentinel-2 imagery. One of the most significant findings is that OSM data proved remarkably accurate, aligning closely with classified land-use maps derived from Sentinel-2 satellite imagery. OSM demonstrated exceptional detail in capturing urban features such as building outlines, road networks, and commercial zones with high spatial fidelity. The validation between Sentinel-2 imagery and OSM derived data strengthened confidence with an overall accuracy of 92.4% and a kappa coefficient of 0.89. Thus, despite its crowdsourced origins, OSM proves that it can be a reliable source for temporal land-use monitoring when properly validated and visualized.

1. Introduction

1.1 OSM data

Open Street Map (OSM) is a globally recognized, open-source, collaborative geographic database that empowers a distributed community of volunteers to contribute, edit, and maintain a wide spectrum of spatial data. These data include critical geographic features such as road networks, building footprints, land use patterns, natural features, and various points of interest. Since its inception in 2004, OSM has grown into one of the most comprehensive and frequently updated sources of volunteered geographic information (VGI), serving as a vital alternative to proprietary geospatial datasets. Its accessibility, openness, and adaptability have led to its widespread adoption in urban planning, navigation systems, disaster management, humanitarian operations, and academic research. This is especially pertinent in regions where authoritative spatial data is unavailable, outdated, or prohibitively expensive to access.

Despite the growing reliability and depth of OSM, concerns about the temporal accuracy, thematic consistency, and spatial completeness of its data persist—particularly in the context of fast-changing urban environments. Urban expansion, infrastructure development, land-use transformation, and construction activities are inherently dynamic processes, and their real time mapping presents a significant challenge. While OSM's community-driven model enables rapid updates, the uneven distribution of contributors and variations in data validation mechanisms mean that not all regions are mapped with the same precision or frequency. Consequently, questions arise about the extent to which OSM accurately captures real-world urban transformations over time and whether it can be relied upon for longitudinal urban studies.

This research addresses these concerns by conducting a detailed spatio-temporal change detection analysis of OSM data, specifically focusing on built-up features such as residential zones, commercial areas, and construction sites. The temporal window of this study spans from 2016 to 2024, a period during which substantial urban growth has occurred in many regions. The primary aim is not only to identify and visualize changes in built-up land use over this time period but also to evaluate the reliability and validity of these changes by comparing them against classified satellite-based data derived from Sentinel-2 imagery. This dual focus allows the study to both document OSM-recorded urban growth and assess the fidelity of that data against independently sourced Earth observation evidence**.**

The utility of OSM for urban planning and scientific analysis has been thoroughly investigated by Braune and Klump, 2014 who explored its quality and usability. They noted that while OSM provides invaluable detail and accessibility, inconsistencies in tagging, geometry precision, and completeness vary significantly across regions, limiting its standalone reliability. While data processing accuracy is crucial, the accessibility and interactivity of these systems have been greatly enhanced by integrating AI-driven chat interfaces. The advent of large language models (LLMs), such as GPT, has fundamentally reshaped how users interact with complex data systems. Dam et al. 2024 have provided a complete survey of LLM-based AI chatbots, detailing the underlying architectures, training paradigms, and practical applications across domains. Their study emphasized the transformative impact of LLMs in knowledge retrieval, summarization, and reasoning. Building on this, Pokhrel et al. 2024 demonstrated a framework combining OpenAI's GPT with LangChain and Streamlit for document summarization and question answering, showcasing how LLMs can serve as intuitive frontends for structured databases and textual repositories. Similarly, Vidivelli et al. 2024 have explored efficiency-driven chatbot development using LangChain, Retrieval-Augmented Generation (RAG), and LLM 7 fusion. Their results affirm that combining these components enables a responsive and context aware interface, capable of synthesizing data from multiple sources in real-time.

Documentation and community-driven knowledge bases such as the LangChain docs, OpenAI Developer Community, and Flask tutorials played a critical role in implementing these functionalities. These sources provide real-world examples, design patterns, and debugging solutions that bridge the gap between theoretical architecture and practical deployment. For OSM-specific spatial information, the OSM Wiki remains a vital reference point, offering detailed insights into tagging conventions, geometry standards, and administrative boundaries (Digital Ocean).

2. Objectives

2.1 To develop a SQL-based spatio-temporal change detection framework for analysing OpenStreetMap (OSM) data, specifically focusing on built-up features such as residential, commercial, and construction-related land uses between the years 2016 and 2024.

2.2 To design and implement a web-based geospatial application using Python Flask that enables interactive visualization of detected changes and integrates a Langchain powered chatbot interface for natural language querying and statistical retrieval.

2.3 To validate the accuracy and completeness of OSM-derived built-up changes through comparative analysis with a supervised Land Use Land Cover (LULC) classification generated from Sentinel-2 satellite imagery.

3. Study area

Luxembourg is a compact, landlocked country in Western Europe, covering just under 2600 km². It stretches from 49°27′ N to 50°11′ N latitude and 5°40′ E to 6°32′ E longitude,

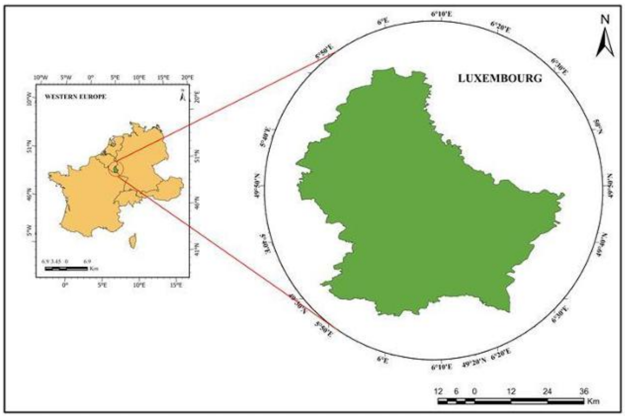

Figure 1 : Location of the study area

bordered by Belgium to the north and west, Germany to the east, and France to the south (figure 1). Despite its small footprint, the nation's terrain is remarkably varied: the Oesling in the north features forested plateaus and hills rising to about 560 m, while the Gutland region in the south descends to around 133 m along the Moselle River, with gently rolling plains that host most of the population and economic activity. A temperate oceanic climate—mild winters, warm summers, and steady rainfall year-round—further shapes how land is cultivated, built upon, and conserved.

4. Data and methodology

The foundation of this research was built upon geographic data sourced from Geofabrik, a widely trusted provider of regional extracts from the global OSM database. The data obtained were in Protocolbuffer Binary Format (PBF), known for compact storage and efficient handling of extensive geographic datasets, which suited the project's large-scale analytical requirements to manage and analyse these spatial datasets effectively, the open source database software PostgreSQL was utilized, along with PostGIS, its spatial extension. This

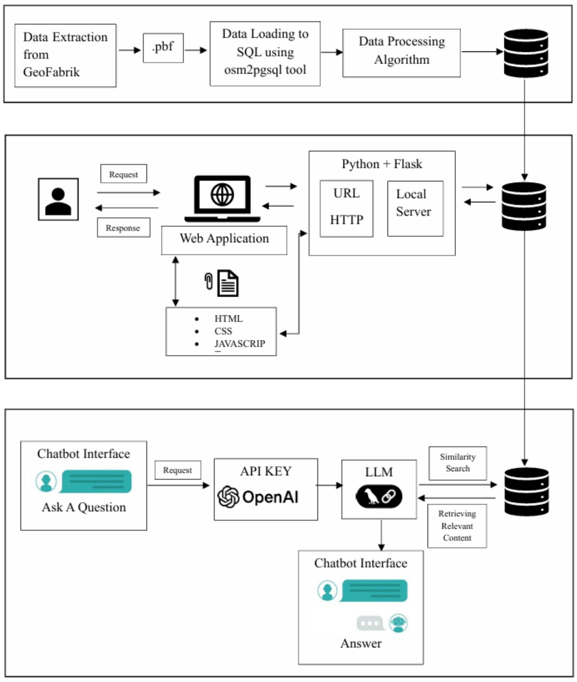

Figure 2 : Methodology

combination was crucial in enabling advanced geographic queries, spatial indexing, and effective data management. To load the downloaded OSM data into this spatial database, a specialized utility called osm2pgsql was employed. This utility simplifies the transformation and integration process of OSM data into structured database tables suitable for spatial analysis. For backend processing and development, the Python programming language (version 3.x) was chosen due to its simplicity, flexibility, and extensive availability of GIS-related libraries. Additionally, Flask, a lightweight and user-friendly Python web framework, was implemented to develop the web application backend. To ensure smooth interaction between Python scripts and PostgreSQL, the psycopg2 library was utilized, facilitating secure and efficient database connections. The frontend part of the web platform was developed using basic web development technologies—HTML5 and CSS3 provided the website's fundamental structure and design, whereas JavaScript enhanced the user experience by enabling interactive elements and dynamic content loading. An integral part of the research was the development of an intelligent chatbot interface. This chatbot utilized the OpenAI API, offering state-of-the-art language model capabilities. Additionally, transformer-based models (such as those available via sentence-transformers) were used to perform semantic similarity searches, enabling contextually relevant answers. These components allowed the chatbot to engage users naturally and responsively, significantly enhancing the interaction experience.

5. Results and discussion

5.1 Phase 1

For this research, the Luxembourg dataset was specifically selected and downloaded in the .osm.pbf format, which is a compressed binary format optimized for handling large volumes of spatial data. To enable efficient processing and spatial analysis, the data was imported into a PostgreSQL database integrated with PostGIS—an extension that adds support for geographic objects to the PostgreSQL database. This data import was carried out using the command-line utility osm2pgsql, which is widely used for loading OSM data into a PostGIS enabled database. The tool not only parsed the raw OSM data but also converted it into spatially structured formats, created geometry columns, and generated spatial indexes. These indexes play a crucial role in accelerating spatial queries and improving the overall responsiveness of the database during geospatial analysis.

Figure 3:Database creation for 2016 and 2024

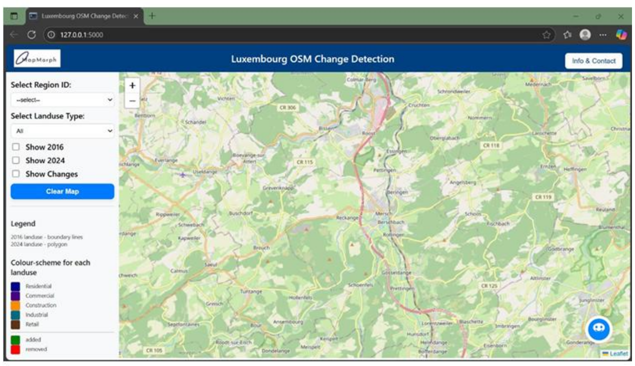

Figure 4: Interface of web application

5.2 Phase 2

A clearly defined color-coded legend helps users quickly distinguish between land-use categories such as residential, commercial, industrial, and construction zones, along with green and red indicators for added and removed areas respectively. Additionally, the platform integrates a statistical chatbot, built using a large language model (LLM), which interacts with the database to provide real-time responses to user queries—like listing region names or summarizing change statistics. This combination of spatial analysis, intelligent querying, and interactive visualization makes the application a robust tool for planners, researchers, and decision-makers studying urban dynamics and land transformation in Luxembourg.

5.3 Phase 3

The chatbot is built using Langchain, a framework designed to harness the capabilities of large language models (such as those developed by OpenAI) and connect them with external data sources like SQL databases. Langchain allows developers to build applications where the LLM can dynamically understand user questions, translate them into database queries, fetch the data, and present the output in human-readable language. In our case, it enables the chatbot to interact with the PostgreSQL/PostGIS database where OpenStreetMap data for different years and regions is stored. The main idea behind this chatbot was to simplify access to statistical insights about land-use changes in Luxembourg.

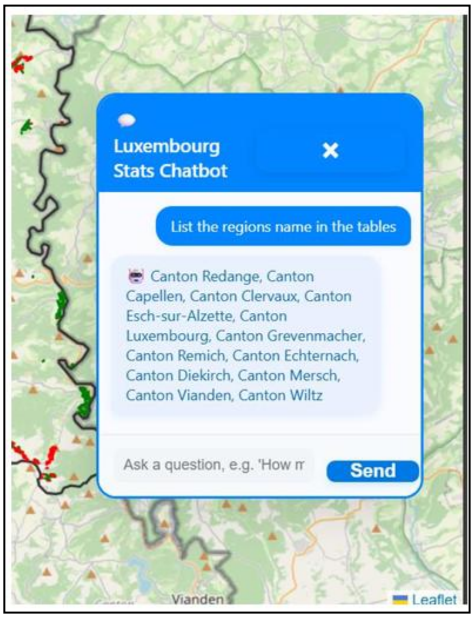

Figure 5: Chatbot

Users can ask questions such as "What are the names of the regions in the dataset?" or "How many residential areas were added in Canton Clervaux?", and the chatbot intelligently interprets the question, runs an appropriate query, and provides a meaningful response. This eliminates the need for technical users to manually write queries or navigate database tables and also empowers non-technical users to extract valuable information effortlessly. From a technical perspective, the Langchain pipeline includes a prompt template specifically designed for our application domain. It connects the LLM with the PostgreSQL database through a SQL chain that executes the correct commands based on the question asked. This approach ensures the chatbot remains accurate, relevant, and context aware for analysing urban structure and land development patterns.

Table 1 presents the total built-up area in 2016, categorized according to OSM (OpenStreetMap) land use tags across various region IDs. The built-up areas are disaggregated into five key land use categories: Commercial, Construction, Industrial, Residential, and Retail, measured in square kilometres. This tabulation allows a detailed understanding of the spatial distribution of built-up categories at a regional level, which is fundamental for analysing urban structure and land development patterns

Table 1 : Area in 2016 using Sentinel 2 image

Region ID	Area Per Buildup Category					Total Buildup Area (sq.km)
	Commercial	Construction	Industrial	Residential	Retail	
1	0.049	0.034	0.211	8.332	-	8.626
2	0.827	0.197	1.97	15.131	0.086	18.211
3	0.43	0.001	0.686	7.46	-	8.577
4	2.213	1.46	9.785	27.85	0.472	41.780
5	3.826	1.021	2.76	30.769	0.41	38.786
6	1.125	0.086	0.236	7.344	-	8.791
7	0.332	0.054	-	5.754	-	6.140
8	0.053	0.036	0.477	4.923	-	5.489
9	0.321	0.353	6.65	0.01	-	7.334
10	0.284	0.125	1.476	8.774	0.269	10.928
11	-	-	0.025	1.8	-	1.825
12	0.018	-	0.181	4.847	-	5.046
						Total Buildup of 2016 =161.533

Table 2 provides a detailed distribution of built-up land across different region IDs for the year 2024, based on OpenStreetMap (OSM) tagging. Similar to the 2016 analysis, the built-up area is categorized into five primary land use types: Commercial, Construction, Industrial, Residential, and Retail, with the total area presented in square kilometers. This classification serves as a key component in identifying the spatial patterns of urban development and monitoring land use dynamics over time.

Table 2 : Area in 2024 using OSM

Region ID	Area Per Buildup Category					Total Buildup Area (sq.km)
	Commercial	Construction	Industrial	Residential	Retail	
1	0.193	0.155	0.43	10.423	0.01	11.211
2	1.044	0.853	2.017	14.787	0.124	18.825
3	0.68	0.315	1.082	10.298	0.108	12.483
4	3.045	1.852	8.884	26.361	0.699	40.841
5	5.718	1.953	2.839	31.665	1.189	43.364
6	1.315	1.315	0.28	10.513	0.038	13.461
7	0.576	0.156	0.62	6.543	0.044	7.939
8	0.26	0.213	0.588	6.785	0.036	7.882
9	0.716	0.367	0.799	10.151	0.052	12.085
10	0.69	0.352	2.13	9.478	0.217	12.867
11	0.054	0.041	0.81	2.03	-	2.935
12	0.198	0.3	0.517	8.413	0.164	9.592
						Total Buildup of 2024 =193.485

A comparison of the built-up areas in 2016 (Table 1) and 2024 (Table 2) reveals a total increase of 31.952 sq.km in the overall built-up land across all regions, rising from 161.533 sq.km in 2016 to 193.485 sq.km in 2024. This translates to an approximate 19.78% increase in built-up area over the 8-year period, highlighting the effects of urban expansion, population growth, and developmental activities observed in the study area.

5.4 Validation

To evaluate the accuracy of the OpenStreetMap (OSM)-derived built-up area data used in this study, a rigorous validation process was conducted using Sentinel-2 satellite imagery for the year 2024. Sentinel-2, with its 10m spatial resolution and multispectral capabilities, provides high-quality Earth observation data suitable for land use and land cover (LULC) classification.

Table 3 : Year wise comparison of area using Sentinel 2 image

Class Name	2016	2024
Build up Area (Sq.km)	129.22	168.37

A supervised classification approach was applied to the imagery, using the Random Forest (RF) algorithm—an ensemble-based machine learning classifier known for its robustness, non parametric nature, and high classification accuracy, particularly in heterogeneous urban environments. Training samples were carefully selected for five land cover classes: Built-up, Vegetation, Water, Barren land, and Agricultural land, based on visual interpretation and prior knowledge. The classification was carried out using Google Earth Engine (or relevant remote sensing software), and the resulting LULC map was validated using a confusion matrix generated from a set of independent ground truth points. The overall classification accuracy achieved was 92.4%, and the kappa coefficient was 0.89, indicating a very high level of agreement between the classified Sentinel-2 image and the actual land cover on the ground.

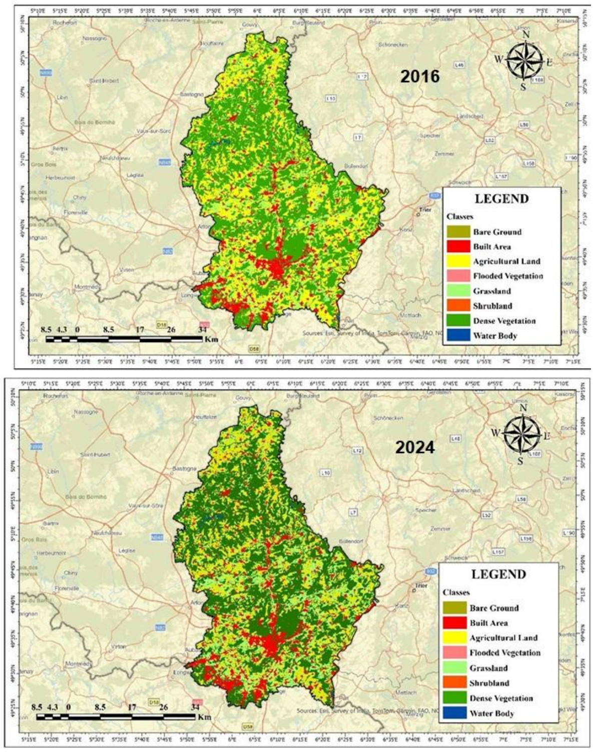

Figure 6: LULC of Luxembourg using Sentinel 2

6. Conclusion

The results of this study bring forward several important insights regarding the use of Open Street Map data for mapping built-up areas and detecting land use change over time. One of the most significant findings is that OSM data proved to be highly accurate, aligning closely with classified land use maps derived from Sentinel-2 satellite imagery. Despite being crowd-sourced, the OSM dataset captured detailed urban features such as building outlines, roads, and commercial zones with impressive precision. In many regions, it even offered better spatial detail than remote sensing methods, which sometimes struggled with spectral confusion in mixed land use zones. The validation process using a Random Forest classifier confirmed the strength of this approach, yielding an overall accuracy of 92.4% and a kappa coefficient of 0.89, which indicates a strong match between satellite-based classification and the real-world land cover. The findings showed that OSM data can be trusted for large-scale spatial analysis, particularly when verified using machine learning and satellite imagery. Additionally, the analysis revealed significant increases in built-up area between 2016 and 2024—most notably in regions 5, 6, and 12—indicating patterns of suburban growth, infrastructure development, and housing expansion, driven by Luxembourg's economic and population dynamics.

Together, these findings confirm that integrating OSM data with satellite imagery and modern web technologies provides a powerful, accurate, and scalable solution for land use monitoring and urban change analysis.

References

Braune, C., Klump, J., 2014: Exploring the Quality and Usability of OpenStreetMap Data. EGU Gen. Assembly, Geophys. Res. Abstracts, 16, EGU2014-6794.

Dam, S.K., Hong, C.S., Qiao, Y., Zhang, C., 2024: A Complete Survey on LLM-based AI Chatbots. arXiv:2401.05442. doi.org/10.48550/arXiv.2401.05442

Pokhrel, S., Ganesan, S., Akther, T., Karunarathne, L., 2024: Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models. J. Inf. Technol. Digit. World, 2024(1), 006. doi.org/10.36548/jitdw.2024.1.006

Vidivelli, S., Manikandan, R., Dharunbalaji, A., 2024: Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, RAG, and Performance-Optimized LLM Fusion. Comput. Mater. Continua, 2024. doi.org/10.32604/cmc.2024.054360

DigitalOcean, n.d.: How to Make a Web Application Using Flask in Python 3. [Online]. Available: https://www.digitalocean.com/community/tutorials/how-to-make-a-web-application-using-flask-in-python-3