Urban Land Use Classification in Metro Manila using Deep Learning

Jonathan Christian F. Aceron [1,2], Gilson Andre M. Narciso [1,2], Rose Anne I. Coronado [1], Lorrize Mae L. Guevarra [1,2], Abdel Jalal D. Sinapilo [1,3], Jeromalyn A. Palma [1] , John Harold B. Tabuzo [1]

[1] Department of Science and Technology - Philippine Institute of Volcanology and Seismology (DOST-PHIVOLCS), Philippines - jcaceron.lupa@gmail.com; gmnarciso.lupa@gmail.com; racoronado.lupa@gmail.com; llguevarra.lupa@gmail.com; ajsinapilo.lupa@gmail.com; jeromalyn.palma@phivolcs.dost.gov.ph; harold.tabuzo@phivolcs.dost.gov.ph

[2] Department of Geodetic Engineering, University of the Philippines - Diliman, Philippines
[3] Artificial Intelligence Program, University of the Philippines - Diliman, Philippines

Keywords: EfficientNetV2, Street View Imagery, image classification

Abstract

Land use maps play a pivotal role in proper land use planning, as they represent not just the physical characteristics of the land but also the various socio-economic activities that it undertakes. Proper land use planning and monitoring ensure that urbanization and development is sustainable. In the Philippines, the Department of Human Settlements and Urban Development (DHSUD) has land use guidelines as part of the Comprehensive Land Use Plan (CLUP) preparation in each city or municipality. While remote sensing technologies have made mapping easier, it can only see building roofs, which makes land use monitoring a tall order. Land use monitoring often requires conventional efforts, including ground validation surveys to account for the rapid changes in land use patterns. This study developed a methodology to rapidly classify urban land use in Metro Manila. The training dataset is composed of images that were scraped from Google Street View from selected points of interest in the region. A neural network named EfficientNetV2 was used for training due to its great accuracy in image classification tasks while being fast and lightweight. The trained model achieved a 60.4% overall accuracy for the testing dataset. The class f-1 score ranges from 0.40 for the Government Office class up to 0.83 for the Parks class. The developed methodology exhibited its capability in rapidly creating land use maps, especially in urban areas, highlighting its potential to be used in urban planning applications and research.

1. Introduction

Land use patterns mutually evolve with the socio-economic and environmental shifts at multiple scales. The intricacy of different factors, including (but not limited to) human decision-making, changes in the land cover, and urbanization, has made land use studies a popular topic over the years. Understanding the complex interaction of these elements leads to the creation and update on the land use maps, which facilitate proper land use planning, management, and monitoring. These maps are essential for strategies and policies that impact sustainable use of land resources.

In the Philippines, the Department of Human Settlements and Urban Development (DHSUD) is the mandated agency in developing guidelines on how to prepare the Comprehensive Land Use Plan (CLUP) for each city or municipality, particularly the standard land use classes have been published (Housing and Land Use Regulatory Board, 2013). Land use maps are crucial in the CLUP development, particularly in the planning process where it is one of the key outputs. The planning body assesses the city/municipality's existing land uses in relation to other relevant data to identify the issues, potential and future development needs, and spatial requirements. Remote sensing imagery has been widely used for rapidly mapping hazards, topography, and green spaces, and has proven effective in supporting the efficient preparation of such plans. However, mapping of the current land use of urban areas has been a challenge due to the limitation set by the orthographic nature of satellite imagery. In addition, characteristics and patterns that differentiate urban land use types from each other are not spectral in nature. Land use mapping of the existing land use is often done on the ground which is often time-consuming and requires a lot of manpower.

Street View Imagery (SVI) are geotagged images acquired from the horizontal plane, usually from a human-level (Biljecki and Ito, 2021). These images provide a different kind of information on the surrounding areas compared to the aerial perspective. There are various SVI sources such as Apple Maps Look Around, Microsoft Bing Streetside, Mapillary, Kartaview, and most notably Google Street View (GSV). Google Street View provides the most coverage and more consistent quality in its images. SVI can provide a faster way of collecting land use data due to its accessibility and abundance, unlike mobile/standard photography from foot surveys, which requires a lot of planning and resources.

Advances in computer vision and machine learning techniques enable various fields of study to utilize the large amount of images available from various SVI data sources (Biljecki and Ito, 2021). Spatial data infrastructure, health, urban perception, transportation, greenery, and walkability are the most frequent topics of related SVI studies. One study leveraged the use of SVI to evaluate the association between Utah's built environment and its health outcomes from 2017 to 2019, which might have been both time- and resource-intensive using traditional data collection methods (Nguyen et al., 2022). Another study used SVI with machine learning techniques to overcome the limitations of collecting pedestrian volume data, which is an important indicator of urban walkability and vitality (Chen et al., 2020). There are also studies which explored the use of deep learning to classify SVI as residential or non-residential (Li, et. al., 2017). Street view imagery can make not just the data collection but also the analysis of urban land use monitoring more efficient.

This study aims to develop a methodology to rapidly extract urban land use using readily available open-source datasets. It aims to do that by performing image classification on street view imagery. The training dataset is composed of images that were acquired from GSV of the selected points of interest in the city. A neural network named EfficientNetV2 was used for training, since it has great accuracy in image classification tasks while being fast and lightweight (Aggarwal et. al., 2023).

2. Materials and Methods

2.1 Study Area

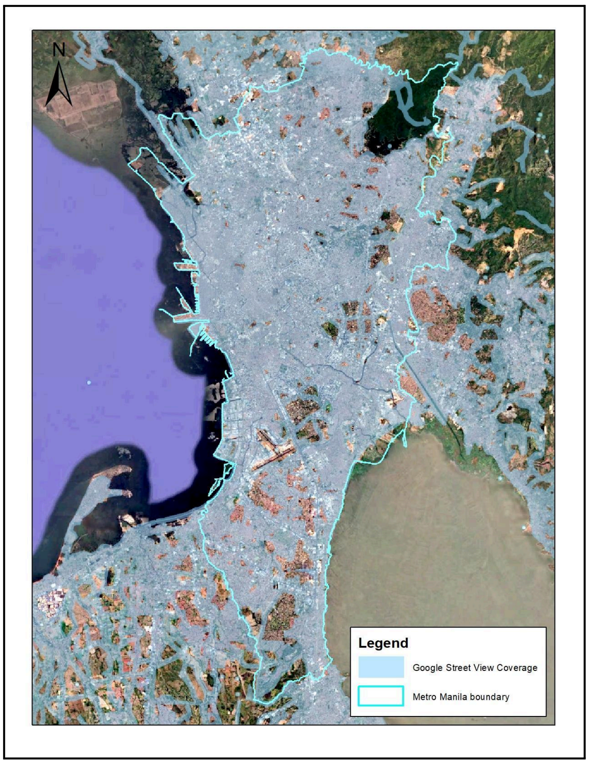

Figure 1. Google Street View coverage of Metro Manila Retrieved from Google Earth Pro.

The study area is Metro Manila, Philippines. It is home to around 14 million people and is the most densely populated urban region in the country (Commission on Population and Development, 2025). Due to its accessibility for SVI, most of the region is covered in GSV as shown in Figure 1.

2.2 Methodology

The methodology has two main parts: the data acquisition and postprocessing, and the training proper and accuracy assessment. Figure 2 shows the entire workflow of the research.

2.2.1 Data Acquisition: An automated web tool based on Python was developed to navigate to randomly selected GSV locations within the study area and extract images. Images were acquired from each location until the minimum number of images for each class was obtained. An example of an extracted image is shown in Figure 3.

Information such as latitude, longitude, and camera orientation was extracted from each image in order to geotag the data. Afterwards, each image was cleaned using a Python script that removes unwanted pixels and retains only the scene to be classified.

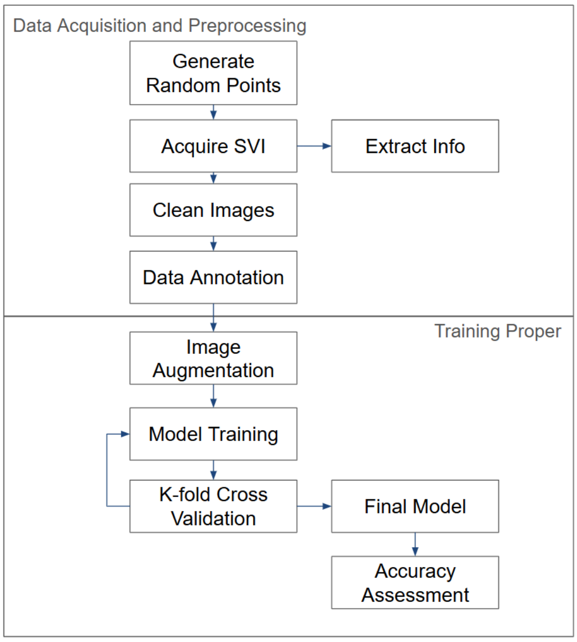

Figure 2. Research Workflow.

Figure 3. Example of a street view image extracted from the web.

2.2.2 Data Annotation: The images are then annotated following the recommended land use classes from CLUP Guidebook Volume 1 and Climate and Disaster Risk Assessment (CDRA) Supplemental Guidelines of Housing and Land Use Regulatory Board (HLURB) and adopted by DHSUD as shown in Table 1. Some urban land use types were further classified for better model accuracy and future studies. For example, the Residential class are separated into Residential and Informal Settlements since each of them have distinct appearances. The same was done on Commercial and Mixed Commercial areas. Mixed Commercial areas are locations with mixed commercial and residential use, usually the lower floors serve as commercial businesses while the higher floors are residential dwellings. The Institutional class was also subclassified to Government Offices, Hospitals, Academic Institutions and Religious Establishments since besides their different appearances, separating them is more appropriate in planning disaster risk response. A total of 10 classes were used.

CLUP Land Use Classification	Land Use Classes in this study
Residential	Residential
	Informal Settlements
Commercial	Commercial
	Mixed Commercial / Residential
Industrial	Industrial
Institutional	Government offices
	Hospitals / Clinics
	Academic Institutions
	Religious Establishments
Parks	Parks

Table 1. Land-use classes used in the study.

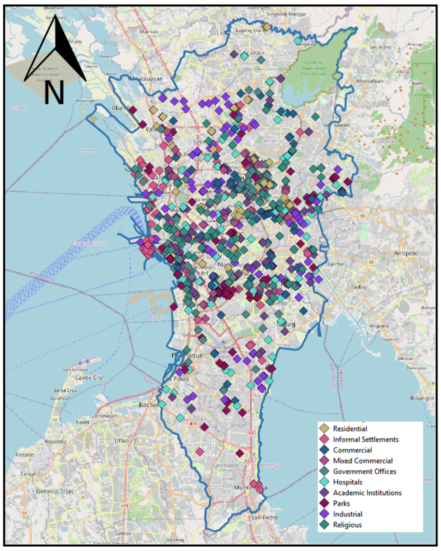

Figure 4. Data Points Collected throughout Metro Manila.

A total of 3000 images were obtained with 250 images per class for the training and validation sets, and 50 images each class for testing. Stratified random sampling was done to distribute the images to train/validate and test sets. The random SVI locations were ensured to be distributed throughout Metro Manila. Figure 5 shows some samples of the annotated dataset.

2.2.3 Model Set-up: The model is implemented through PyTorch, which is a Python based, open-source machine learning framework. The EfficientNet Model was obtained through the timm library. EfficientNet was selected since compared to other image classification models, it is lightweight without sacrificing accuracy. The V2-S version was chosen since it has big improvements of up to 11x faster while being 6x lesser in size from the original EfficientNet (Tan and Le, 2021). The S version was chosen since it is the more compact version and can be used on more machines.

2.2.4 Training Proper: During training, each image is augmented by reducing the image size to 512 by 512 pixels and some images are randomly flipped. The CrossEntropyLoss was used along with the Adam Optimizer. A k-fold cross validation was also implemented to ensure that the best model is not biased towards a particular set of data. The selection of images for training and validation set during the training proper is randomized. The GPU used is a RTX 4070 Mobile GPU. A batch size of 16 was found to be the limit for the GPU.

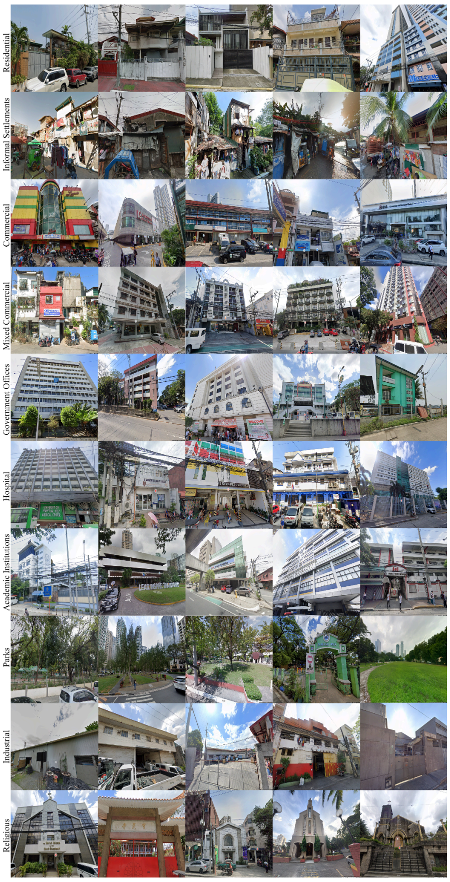

Figure 5. Examples of SVI images collected.

2.2.5 Accuracy Assessment: The best model will then be used to perform inference on the testing set which has 50 images per class. A confusion matrix and other metrics was then generated from the result.

3. Results and Discussion

3.1 Training Accuracy

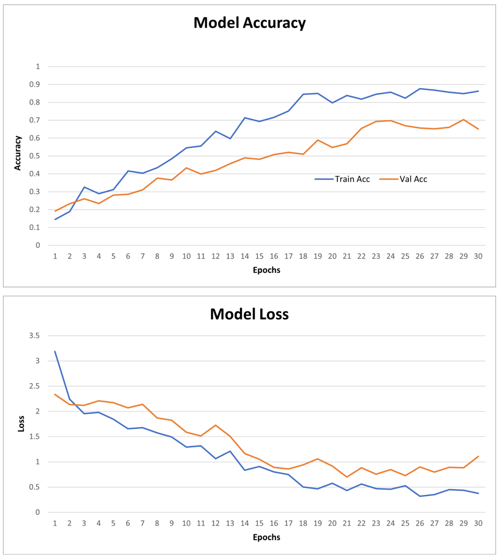

Figure 6. Best Model Accuracy and Loss for 30 epochs.

The training was done for K=2 folds, and after running for 30 epochs, the training accuracy and loss stabilizes to around 82% for the accuracy and 0.4 loss value. Both models exhibit similar accuracy, only deviating by 2-3% throughout the training. The best model's accuracy and loss is shown in Figure 6. Meanwhile, the validation accuracy and loss throughout the training diverges with the training set, suggesting a slight overfitting on the training data.

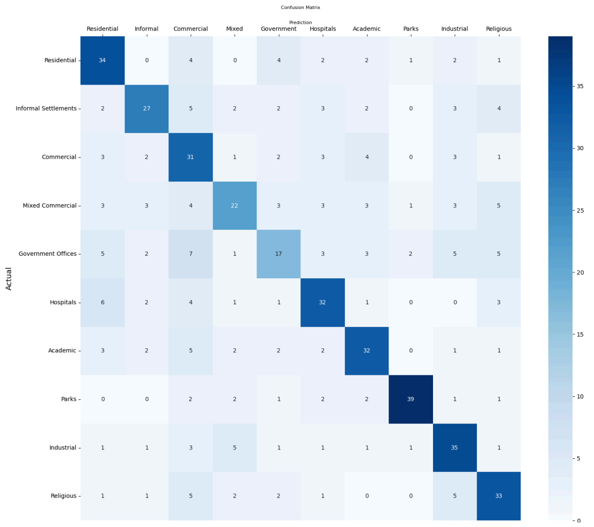

Figure 7. Confusion Matrix for the Test Dataset.

Table 2 shows the accuracy when the best model is applied to the testing set. Parks and Informal Settlements exhibited the best precision with 0.89 and 0.68 respectively, unlike the Commercial and Mixed Commercial classes with 0.44 and 0.58. Meanwhile, the Parks and Industrial class exhibited the best recall with 0.78 and 0.70 compared to the worst in Government and Mixed Commercial with 0.34 and 0.44. Overall, the Parks class exhibited the highest balanced accuracy with 0.83 f-1 score followed by the Industrial class with 0.65, and then third is the Academic class with 0.64. The least accurate classification is the Government Office class with 0.4 f-1 score. The overall accuracy of the model is 60.4%.

Class	Precision	Recall	f-1 score
Residential	0.59	0.68	0.63
Informal Settlements	0.68	0.54	0.60
Commercial	0.44	0.62	0.52
Mixed Commercial	0.58	0.44	0.50
Government Offices	0.49	0.34	0.40
Hospitals / Clinics	0.62	0.64	0.63
Academic Institutions	0.64	0.64	0.64
Parks	0.89	0.78	0.83
Industrial	0.60	0.70	0.65
Religious	0.60	0.66	0.63

Table 2. Image Classification Accuracy.

3.2 Class Prediction Analysis

Figure 8 shows some examples of the prediction with the first two columns showing the correct predictions and the last two columns showing the incorrect ones. Upon investigating the prediction, it is understandable that Parks exhibited the greatest accuracy since it has the most consistent appearance of all the classes. Parks have notable abundance in greenery and have less structures. It is also consistent in architecture since it comprises mostly open spaces. Some classes get misclassified to Parks when there is an abundance of greenery as shown in Figure 8 - R1, C3. The Industrial class showed great recall and moderate f-1 score, as their appearances are mostly monochromic and walled up structures. Misclassifications occur when the Industrial images become abundant in color and in irregular structures (Figure 8 - R9, C4 - C5).

The Religious class has three groups: Catholic/Christian churches, Mosques, and Chinese Temples. Chinese temples and Mosques have great accuracy because of their distinct gates and domes. Meanwhile, the Churches vary greatly in architecture depending on the denomination and age. They are mostly misclassified with Commercial areas and Industrial. An example is Figure 8 - R9, C4 in which a church is hidden behind a wall.

Hospitals have a distinct white and grey appearance. They are often multi-floor in nature and with regularly patterned windows. Because of this, they are often misclassified with Residential and Commercial Areas with similar height and pattern. They are also often misclassified with Religious churches because of the color.

Academic Institutions have two main appearances: first are two or three level buildings with surrounding walls which are commonly for elementary and high school institutions. This type of academic buildings vary in color as some are yellow or blue and some are cream, white, or gray. The second type are multi-level buildings which are mostly panchromatic in appearance. This type of Academic Institutions are mostly misclassified as Commercial establishments.

Commercial establishments show a low f-1 score because of its highly varying appearance. Upon investigating the images, some single unit establishments such as fast food restaurants and distinct brand stores show a few variations in color and features while some commercial establishments such as strip malls and stores have high variations in color. They are often misclassified with Academic, Hospitals, and Residential classes - mostly multi-level and differently colored structures.

Government Offices exhibited the least recall and least f-1 score, which is understandable since their appearance varies significantly depending on their function. They are also age-dependent, newer government buildings show a modern design compared to old ones. They often get misclassified to Commercial establishments.

Separating the Residential from Informal was the right decision, as shown by the fact that they have very few confusions with each other. The Informal class is abundant in roofing sheets and makeshift materials. They are often clustered together making a highly heterogenous image. The Residential class however, is moderately confused with both Commercial and Government classes. The Residential class contains both low rise and high rise dwellings, which was the probable source of confusion. The Mixed commercial class also exhibited a low f-1 score, and investigating the images show that it is also greatly varying in appearance as some areas are high-rise condos with stores at the ground floor and some areas are general two-level Mixed residential buildings.

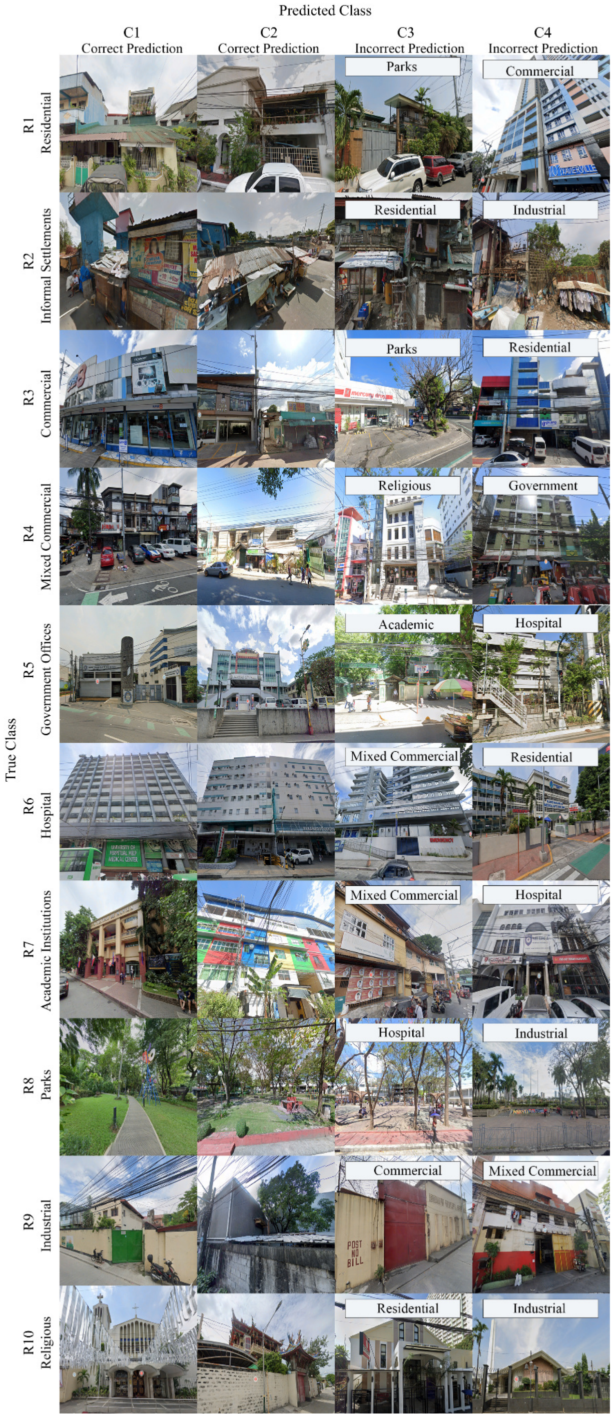

Figure 8. Example predictions of the best model.

3.3 Image classification applications

The classification results highlight the potential of the methodology in land use classification. By generating points where there is streetview coverage, one could rapidly generate land use data without going there physically. This is especially useful in very dense regions such as Metro Manila where there are a lot of buildings which need to be mapped. The results of this study is not limited to the use of SVI, as photos taken from mobile phones and cameras could substitute for the training dataset and be used in inference, with the drawback that is the time and logistics. However, the current results still cannot replace the traditional methods because of its limited availability and frequency. It can best serve as a complement to the physical surveys. By doing automated land use classification first, the areas that need to be visited could be reduced by a lot.

Regarding its accuracy, further subclassifying the classes based on the exact structure will increase the accuracy of the model, as shown in the correctly and incorrectly classified images. However, this requires significantly more training images. Even though the current training dataset is relatively small compared to the number of available images in the region, the image annotation consumed a significant time in the research.

4. Conclusion

This study used Street View Imagery, along with deep learning, particularly image classification, to classify images based on land use type. The developed methodology exhibited its capability in extracting land use information, especially in certain classes where it exhibited high accuracy. Since the methodology from data gathering to inference is automated, the methodology has potential in the rapid mapping of land use type once the model has been created. The study also shows the potential of using image classification in highly complicated tasks.

The researchers have some recommendations, particularly on the implementation of the deep learning model. An Explainable AI (XAI) Technique such as GradCAM could be used to generate heatmaps and show the decision making of the network in making its prediction. Along with increasing the number of images, studying the heatmaps could lead to better classification approaches.

The viewing angle of the structures could also contribute to the result since the image varies greatly in viewing angle depending on the location of the camera. This warrants further testing and the use of a pre-trained foundation model could be beneficial since it is trained on billions of images with different angles. The abundance of clutter in the images should also be addressed, since in most images there are cars, people, and electric wiring often within the view of the structures.

Acknowledgements

This research was done through the Department of Science and Technology (DOST) Grants-in-Aid Project entitled: Spatio-Temporal Land Cover Mapping for Risk Assessments in Land Use Planning through AI (LUPA Project) which is under the Accelerated Earthquake Multi-hazards Mapping and Risk Assessment Program of the Philippines (ACER Program) monitored by the Philippine Council for Industry, Energy & Emerging Technology (DOST-PCIEERD) and implemented by the Philippine Institute of Volcanology and Seismology (DOST-PHIVOLCS).

References

Aggarwal, S., Sahoo, A.K., Bansal, C., & Sarangi, P.K., 2023: Image classification using deep learning: A Comparative Study of VGG-16, InceptionV3 and EfficientNet B7 Models. *2023 3rd International Conference on Advance Computing and*

Innovative Technologies in Engineering (ICACITE) (pp. 1728-1732). IEEE. https://doi.org/10.1109/ICACITE57410.2023.10183255

Biljecki, F., & Ito, K., 2021: Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning, 215*, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Housing and Land Use Regulatory Board, 2013: CLUP Guidebook: A Guide to Comprehensive Land Use Plan Preparation. Volume 1 - The Planning Process.

Li, X., Zhang, C., & Li, W., 2017. Building block level urban land-use information retrieval based on Google Street View images. *GIScience & Remote Sensing, 54(6)*, 819–835. https://doi.org/10.1080/15481603.2017.1338389

Long C., Yi L., Qiang S., Yu Y., Ruoyu W., & Ye L., 2020. Estimating pedestrian volume using Street View images: A large-scale validation test. *Computers, Environment and Urban Systems*, *81*, 101481. https://doi.org/10.1016/j.compenvurbsys.2020.101481.

Nguyen, Q.C., Belnap, T., Dwivedi, P., Deligani, A.H.N., Kumar, A., Li, D., Whitaker, R., Keralis, J., Mane, H., Yue, X., Nguyen, T. T., Tasdizen, T., & Brunisholz, K. D., 2022. Google Street View Images as Predictors of Patient Health Outcomes, 2017–2019. *Big Data and Cognitive Computing, 6(1)*, 15. https://doi.org/10.3390/bdcc6010015

Tan, M., & Le, Q. V., 2021: EfficientNetV2: Smaller Models and Faster Training (arXiv:2104.00298). arXiv. https://doi.org/10.48550/arXiv.2104.00298