Deep Learning for Built-up Fractional Mapping using Sentinel-2 images: A Case Study in Delhi, India

Abhishek Rawat *, Prasun Kumar Gupta *, Claudio Persello *

Posted Date: 26 January 2024

doi: 10.20944/preprints202401.1879.v1

Keywords: fractional mapping; built-up; deep learning; Global Human Settlement Layer; Sustainability

Article

Deep Learning for Built-Up Fractional Mapping Using Sentinel-2 Images: A Case Study in Delhi, India

Abhishek Rawat [1,2], Prasun Kumar Gupta [1] and Claudio Persello [2,*]

[1] Indian Institute of Remote Sensing, ISRO, 4 Kalidas Road, Dehradun 248001, India; abhishekrawat426@gmail.com (A.R.); prasun@iirs.gov.in (P.K.G.)

[2] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands; c.persello@utwente.nl (C.P.)

* Correspondence: c.persello@utwente.nl (C.P.)

Abstract: In our increasingly urbanized world, precise and up-to-date maps of human settlements are essential for sustainable urban development policies. The availability of open-access Sentinel-2 data from the Copernicus program presents an opportunity to create a comprehensive global map of human settlements, offering a detailed view of built areas on a large scale. This study estimates large-scale built-up fractions using encoder-decoder deep learning architectures like U-net, Res-U-net, and Attention-U-net in the large and complex urban area of Delhi, India. Openly available datasets like Open Street Map (OSM) and Microsoft building footprint datasets are used to derive built-up fractions at 10×10m resolution cells for over 34,000 km^2. Our results show that Attention-U-net with the Huber loss function performs the best in different built-up densities (i.e., urban, semi-urban or rural) with an R^2 score of 0.631, while Res-U-net and U-net obtained an R^2 score of 0.623 and 0.612, respectively. The investigated networks significantly improve the accuracy over the latest Global Human Settlement Layer product (GHSL-S2), which uses a deep-CNN and reaches an R^2 of 0.387 in our case study area. The result of this study yields a valuable spatial layer for examining the spatial distribution of human settlements across the entire spectrum from rural to urban areas.

Keywords: fractional mapping; built-up; deep learning; Global Human Settlement Layer; sustainability

1. Introduction

The rapid expansion of cities worldwide presents a challenge discussed among "various research communities", including sustainability, social science, and remote sensing [1]. Urbanization significantly impacts climate at different scales and with the need and want for more and more infrastructure to support the growing populations especially in developing countries where still construction is done using concrete and other heat-absorbing elements rather than sustainable elements, the Urban Heat Island (UHI) effect is very profound [2]. Urban populations in Asia and Africa are projected to double by 2050, with limited data available on settlement patterns, particularly in low and middle-income countries [3–5]. The dynamic transition between rural and urban areas morphology further complicates the evolving structure of cities and regions, challenging mapping, and monitoring efforts of built-up in those regions [4,6].

Long-term missions such as Landsat, a part of the NASA/USGS joint program that started in 1972, and Sentinel, an Earth Observation mission of Copernicus under the European Space Agency (ESA); comprising an array of medium to high-resolution satellites which have offered the opportunity to analyse spatial and temporal urban patterns using image analysis technology to interpret surface features seen from satellites [7,8]. Geographic Information System (GIS) software aids in processing remote sensing data to identify objects and patterns on the ground, contributing to meaningful insights. However, interpreting raster images from remote sensing satellites presents challenges due to mixed pixels caused by real-world landscape variability [9]. These mixed pixels

 doi:10.20944/preprints202401.1879.v1

contain valuable information but also introduce errors in the analysis. Overcoming these errors requires grasping spatial and spectral attributes and advanced algorithms. Spatial attributes influence pixel composition, impacting the occurrence of mixed pixels in the raster image.

Researchers have recently investigated various machine learning strategies for built-up extraction, from conventional classifiers to sophisticated deep learning models. Support Vector Machines (SVM) [10] and Random Forests are popular classification techniques used for categorizing individual pixels according to spectral properties. These techniques have effectively distinguished between populated and unpopulated areas [11,12]. However, with the introduction of object-based image analysis (OBIA), researchers have broadened the analysis to incorporate spatial and contextual information. Benz et al. The authors in [13] utilized OBIA to analyse remote sensing data, incorporating fuzzy logic for GIS-ready information extraction. Multi-scale object-based techniques for image segmentation have substantially improved built-up extraction [14]; however, they are likely to result in the problems of "over segmentation" or "under segmentation" when segmented built-up split apart compared to the ground truth or either mix in with background objects respectively [15].

Traditional machine-learning techniques have greatly enhanced building extraction, although they still substantially depend on hand-crafted features [16]. Additionally, the generated features are manually modified to correspond with the training dataset's building distribution pattern, which may not apply to other datasets, particularly those from other satellites and study sites [17]. The limited transferability of shallow machine learning algorithms with fewer neural layers and complexity restricts their use in automated analysis. Conversely, Deep Convolutional Neural Network (CNNs) enable direct learning of spatial patterns from raw data without needing feature engineering [18]. CNN is a Deep Learning method that can recognize distinct objects and attributes in an input picture and distinguish between them by assigning weights and biases that may be learned. It requires a lot less pre-processing than other classification techniques. While filters are manually designed in a more traditional way i.e., hand-engineered, CNN may learn these filters and attributes given sufficient training data. By automatically deriving hierarchical representations of characteristics from raw data, CNNs have displayed extraordinary performance.

Fully Convolutional Networks (FCN) [19] and U-Net [20], two CNN-based techniques, have demonstrated impressive accuracy in segmenting and defining built-up regions at the pixel level. Yuan et al. [21] introduced an FCN architecture for building extraction from aerial imagery. They accurately and efficiently extracted built-up zones by training the FCN on annotated data. Their work demonstrated how deep learning may be helpful for urban planning and monitoring tools like land resource management, urban sprawl monitoring, cadastral mapping, land use, and land cover, making it easier to extract built-up regions from aerial images automatically. A multi-constraint FCN was proposed by G. Wu et al. [22] to extract buildings from aerial photographs. Sherrah [23] built an FCN without a max pool layer to maintain the precise ground details from the original image and reduce the data loss in pooling.

An adaptation of FCN which was modified to work with fewer inputs using the usual contracting network and a symmetric expanding network where max-pooling was replayed by upsampling operators, U-Net for Land segmentation, was first developed by Seferbekov et al. [24], who mainly focused on high-resolution satellite imagery. Their study confirmed the U-Net architecture's efficiency in defining land cover regions precisely. A semantic segmentation method based on U-Net was suggested by Francini et al. [25] to detect and categorize built-up regions in high-resolution satellite data. Their investigation demonstrated U-Net's capability for producing accurate and trustworthy outcomes for earthquake risk zones in southern Italy.

Y. Wu et al. [26] use attention-based CNN for delineating built-up areas using TerraSAR-X SAR imagery where attention block (i.e., a combination of convolution and activation layers intended to highlight only the relevant activations during training by giving higher or large weights to relevant parts of the image and less relevant parts get small weights) helped minimize the higher and lower false alarm rate caused by weighted cross-entropy loss. A densely linked dual-attention network with a multiscale context was suggested by Chen et al. [27] to extract buildings at the block level. The recommended method divides SAR into multi-scale blocks that partially overlap, utilizing several

 doi:10.20944/preprints202401.1879.v1

size grids and their multiple-step offsets. The feature representation and classification of SAR blocks are carried out by the lightweight network built by fusing dense layers and double attention. Ultimately, it combines predictions from several blocks using pixel-level multi-label voting to recognize built-up zones accurately.

All the studies above have exploited high spatial (Very High Resolution Multi-spectral) [28,29] or sometimes a fusion [30,31] of it with SAR or Hyper-spectral datasets, giving us a classified image with resolution justifiable when it comes to classifying buildings (i.e., edge detection of buildings or built-up in general to its surroundings is more distinguishable with very high-resolution imagery or sub-meter resolution as each pixel even in compact settlements can detect the edge or boundary of the built-up). However, most of these high spatial and spectral resolution products come from commercial satellites, which are not always feasible monetarily for large scales. Researchers have transitioned to using Sentinel-2 data, but its 10-meter resolution remains inadequate for precise mapping [32]. To address this limitation, they have turned to mixed pixel regression, specifically fractional cover analysis [33]. This approach offers a more nuanced understanding of the distribution and density of built-up areas by using fuzzy logic rather than pure pixel classification, benefiting urban planning and environmental assessments.

Using a global composite of Sentinel-2 data, Corbane et al. [34] propose a deep learning-based architecture for fully automated extraction of fractional built-up (probabilistic 0 to 100 distribution of built-up confidence) with 10-meter spatial resolution at the global scale. Using transfer learning, a multi-neuron modeling technique based on a straightforward convolution neural network architecture is developed for generating fractional built-up cover.

However, the methods of Corbane et al. struggled to fully harness the available spatial and spectral information due to limitations of the Deep-CNN used in the study, resulting in inaccuracy in developing countries like India [34]. The novelty of this research lies in using more complex CNN-based semantic segmentation models like U-net [20], Res-U-net [35], and Attention-U-net [36] to generate a large-scale fractional built-up cover as opposed to traditional binary maps. The research also explores different open-source reference built-up labels and assesses them based on their accuracy and reliability for generating reference input ground truth mask datasets for deep learning models.

2. Materials and Methods

2.1. Study Area and Datasets

India's capital, Delhi (NCT), and the National Capital Region (NCR) provide a sizable study area for built-up extraction, around 34,000 square kilometers (as seen in **Error! Reference source not found.**), which is selected to improve heterogeneity in the data, providing us with: 1) "urban" Delhi center, (2) "semi-urban" large satellite towns like Noida, Greater Noida, Meerut (in Uttar Pradesh), Gurugram, Rohtak, & Panipat (in Haryana); and (3) "rural" areas of Haryana and Uttar Pradesh states. This offers an attractive environment for researching built-up regions and their dynamics due to their historical significance, extensive cultural history, and fast urbanization. Understanding this urban-rural structure and morphological patterns will help better comprehend other cities. This information may inform initiatives for sustainable development, policy creation, and urban

 doi:10.20944/preprints202401.1879.v1

planning strategies to address the problems encountered by India's increasingly urbanizing cities.

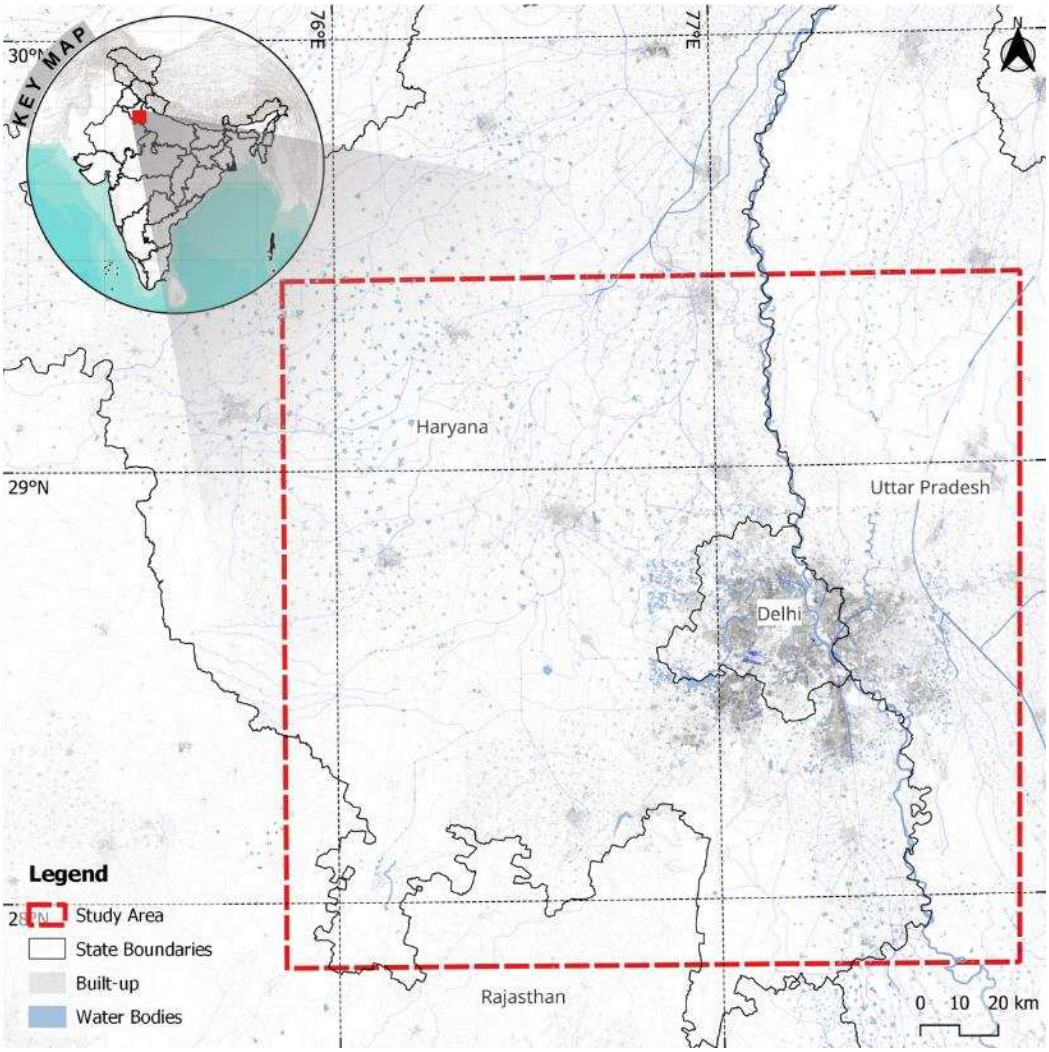

Figure 1. Study area of Delhi NCR, India, spanning 34,000 km^2 (represented with the red dashed box). Grey areas represent built-up (derived from OSM and Microsoft datasets), and blue areas represent water bodies.

2.2. Datasets

The datasets used in this study can be categorized into three parts: (1) input multi-spectral imagery; (2) open-source built-up datasets for generating reference built-up labels; (3) high-resolution multi-spectral imagery. For the input multi-spectral imagery, the Sentinel-2a MSI Surface Reflectance dataset (Surface Reflectance (SR) Sentinel-2 images that have been atmospherically corrected are provided by the Level-2A product which includes corrections for atmospheric corrections for aerosol particle absorption and scattering from December 2018 to the present) was used with a mean image composite of the year 2021-22 from Google Earth Engine (GEE) [35] for cloud-free imagery. The final composite had a spatial resolution of 10 meters and comprised four bands (Band 2: blue, Band 3: green, Band 4: red, and Band 8: near-infrared) which provided the best open-access input imagery at the time of this study.

A crucial issue with any deep learning architecture is the volume of data used for training required to alter the network variables appropriately. A sizable collection of open access and free datasets describe built-up regions in varying degrees of depth, completeness, consistency, and correctness. The most comprehensive datasets characterizing built-up areas were gathered from public sources to create a consistent, thorough, and precise delineation of the study area for generating reference built-up labels. These datasets include Open Street Map (OSM), Global Human

 doi:10.20944/preprints202401.1879.v1

Settlement Layer (GHSL_BU), Facebook high-resolution settlement dataset (FB_HRS), and Microsoft building footprints (MS_BFP).

OSM dataset was downloaded from the Geofabrik website [36], with buildings, roads, and railways considered under built-up in accordance with INSPIRE guidelines [37]. According to Corbane et al. [38], GHSL_BU was created by automatically classifying 2014 Landsat 30 m resolution data. Symbolic Machine Learning (SML) classifier that automatically makes logical rules tying the image dataset to existing high-abstraction semantic tiles utilized for training [39] is the cornerstone of the method for mapping populated areas using global Landsat data. The product has a spatial resolution of 30 m. Despite robust worldwide performance displaying built-up areas, GHSL_BU struggles with under-fitting difficulties in sparsely populated areas, notably in rural Asian settings.

Settlement grids created by Meta (previously Facebook) [40] of high spatial resolution are the source of the FB_HRS data. The "Data for Good" Facebook program, which promotes global humanitarian activities, made the dataset open access [41]. The population zones in FB_HRS were automatically identified by a CNN classifier using sub-meter resolution optical images and high-resolution open-source data for training from OSM [42]. Although FB_HRS data is far spread out in the study area, it represents population zones rather than built-up, and so it has not been used in this study.

Microsoft's Open Source CNTK Unified Toolkit was used to retrieve the data automatically. MS_BFP was generated by performing RefineNet up-sampling layers on Bing images that contain VHR satellite and aerial sensors using CNTK and ResNet34 [43]. MS_BFP dataset is provided in vector format at 1:10000 scale, enabling 1 x 1 m rasterization used in this study, which was then aggregated to 10 x 10 m resolution. The countries where information was available when the MS_BFP data was gathered were the United States, Canada, Africa, the Caribbean, Central Asia, Europe, the Middle East, South America, and South Asia.

2.2. Methodology

The primary objective of this study is to design a deep-learning-based strategy to map built-up fractions using Sentinel-2 data and for that, the study is divided into two sections: data preparation and deep learning modeling.

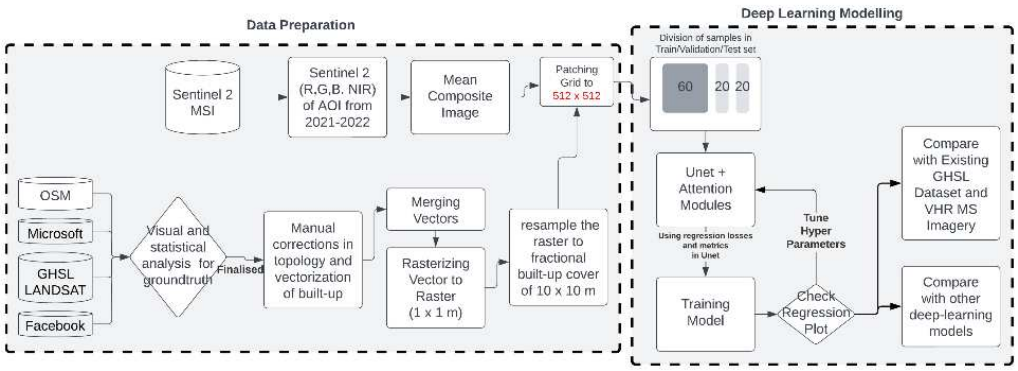

Figure 2. Methodology divided into two parts a) data preparation comprising of data collection, filtering, pre-processing, and patching for modelling; b) deep learning modelling where the models are trained and tested.

The first part, data preparation, involves exploring the input dataset, analyzing it quantitatively and qualitatively and then preparing the dataset for training the deep learning model (hereafter referred to as reference built-up labels). A filtering process was implemented to choose the best data locally available for the input dataset. Out of the datasets collected to generate reference built-up labels dataset at 10 m resolution, Microsoft building footprint (MS_BFP) and Open Street map (OSM) were given precedence first because of better coverage and precision compared to other datasets for

 doi:10.20944/preprints202401.1879.v1

generating reference built-up labels. The Facebook High-Resolution Settlement (FB_HRS) dataset and Global Human Settlement Layers Built-Up Grid (GHSL_BU_2016) had the least accurate representation of urban and rural areas. On the contrary, the Sentinel-2 image composite [44] came pre-processed out of the box through Google Earth Engine [35].

According to INSPIRES's definition of built-up, "any surface above ground level used for any purpose" [37]; buildings, roads, and railways [45–47] data was taken from OSM and building footprint was taken from MS_BFP followed by manual corrections in many urban and some rural areas, which resulted in an accurate dataset of built-up when evaluated against Google Earth's Satellite data.

OSM's "Roads" and "Railways" features were transformed into polygons using QGIS's buffer tool. Various widths (1m, 2m, and 3m) were experimented with, as shown in Figure 3. 1 meter caused uneven pixelation, 2 meters had gaps, but 3 meters worked best for 10m resampling. Wider widths would have led to overestimation in the model prediction.

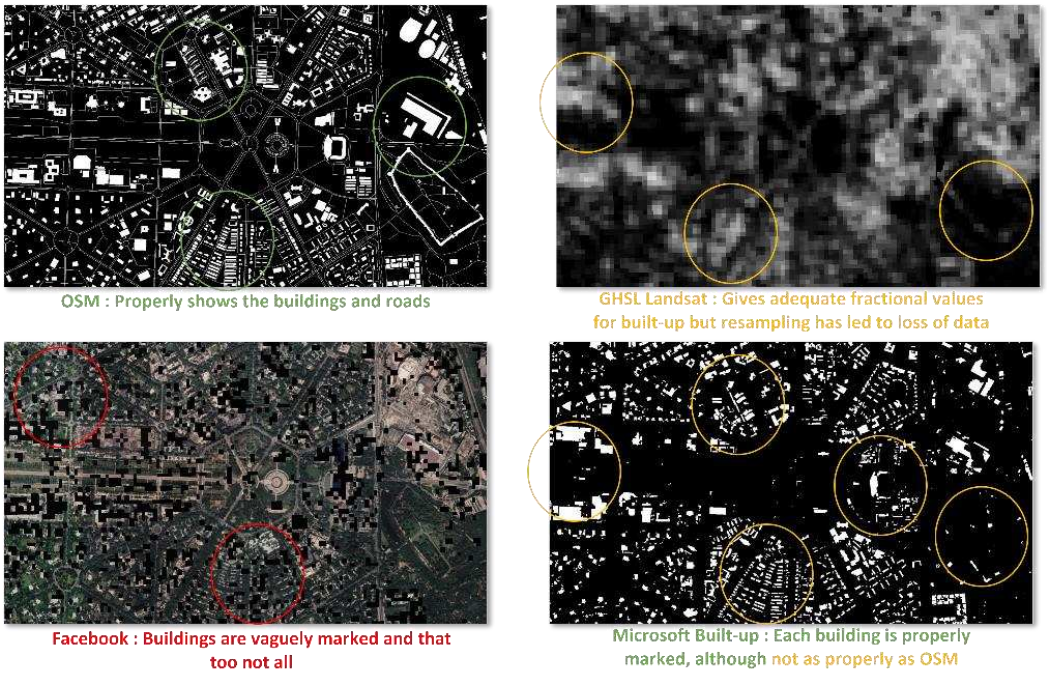

Figure 3. Urban Setting to compare datasets; Green ring = good quality data, Yellow ring = mediocre quality data, and Red ring = data which is not suitable.

 doi:10.20944/preprints202401.1879.v1

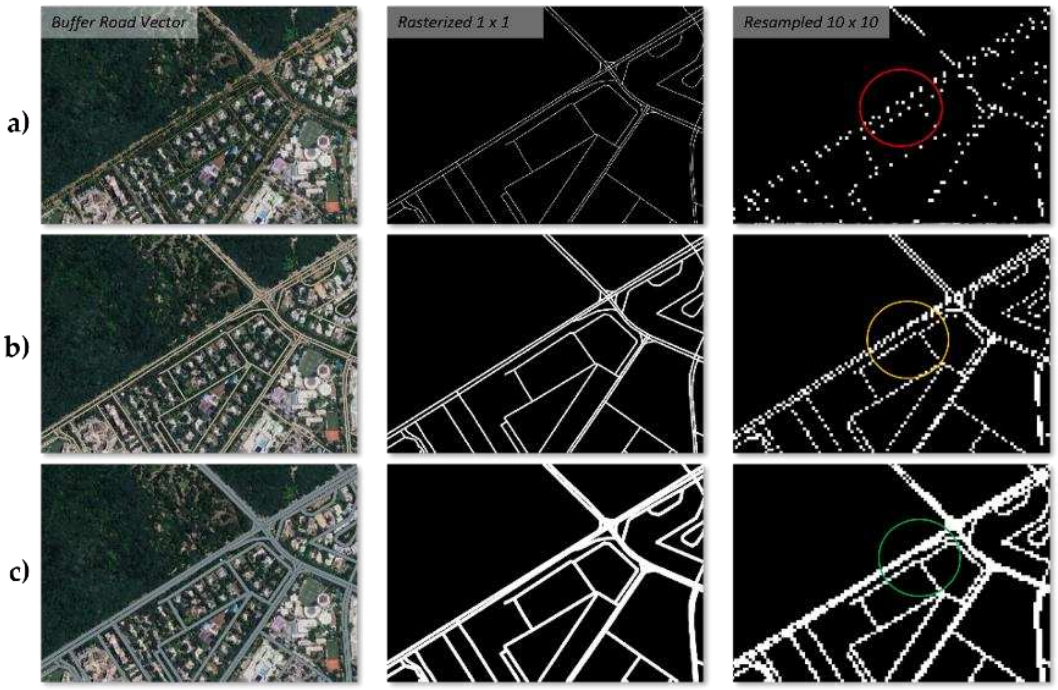

Figure 4. Buffer experimentation: (a) Buffer: 1 m produces pixelated road, not covering properly (top) (b) Buffer: 2 m produces pixelated road, still it is not a continuous stretch of pixels for the road (middle) (c) Buffer: 3 m produces pixelated road, properly covering the road with some gaps (bottom). Source: Author.

After that, all the polygons were merged in QGIS, rasterized to 1 meter, and resampled to 10 meters using GDAL tools [48]. Both the input datasets (image and generated reference built-up labels) were then patched at 512 × 512-pixel level, segregated according to the built-up densities in each patch (high comprising of top 25th percentile, medium comprising of average to 75th percentile and rest coming in low built-up densities) which were then distributed equally into training, validation and testing dataset having a ratio of 60%, 20% and 20% respectively for homogeneity during training and testing.

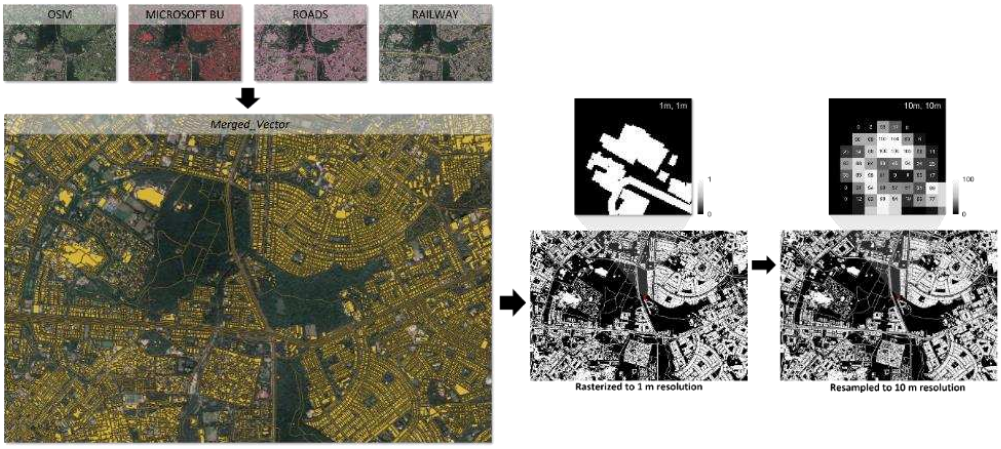

Figure 5. Merging OSM datasets and MS_BFP (left), rasterizing at 1 x 1 m resolution (center), and resampling to 10 x 10 m (right). Source: Author.

The second part explores deep learning architectures, validation criteria, and prediction methods to predict fractional values from the input dataset.

$$z_{i,j} = b + \sum_{k'=1}^{c} w_{k'} \times x_{i,j,k'} \quad (1)$$

 doi:10.20944/preprints202401.1879.v1

$$y_{i,j} = \frac{1}{1+e^{-z_{i,j}}} \quad (2)$$

The sigmoid function (as seen in the equation above) is adopted for prediction at the output layer since binary categorization of building and non-building is needed: The weights and bias are indicated by the symbols $w \in R^c$ and $b \in R^1$. Prediction $y_{i,j}$ can only be made between [0, 1], which can be utilized further in the models adopted in the research to get a fractional built-up output with a value between 0 and 1 with 0 denoting no presence of built-up and 1 denoting 100% presence of built-up in a pixel of the input imagery.

Sigmoid was incorporated at the output layer of the three encoder-decoder models used in the study: (1) U-net [20], modified according to Francini et al.[25] with no dropout layer and a kernel size of 3 × 3; (2) Res-U-net [49], modified according to Yi et al. [50], where one convolution block out of a pair is replaced by a residual block in both encoder and decoder path increasing layer depth and simplifying optimization [51]; and Attention-U-net [52], modified from Y. Wu et al. [26], where the attention gate is attached to the skip connection at the decoder part of the U-net, provides added information at the end of the decoder path by optimizing weightage at the skip connection and limiting the impact of outliers or noise from the input.

The assessment criteria of the model predictions are based on the R^2 and RMSE scores as the models are modified for regression rather than classification by providing an output in the form of probability (using the sigmoid activation function at output), which results in 0 to 1 value which are interpreted as 0 being no built-up and 1 being complete (100%) built-up coverage in the sentinel-2 pixel of 10 × 10 meter.

3. Results

3.1. Training phase of models

3.1.1 Hyper-parameters tuning

In order to get the best possible gradient descent curve (i.e., the fit of the model for the problem), hyper-parameter tuning was done using the Halving grid search technique [53] with the following parameters for each category for all three models used in this study:

The result of halving grid search (as seen in Table 1) was that the best activation, optimizer, loss, and metrics combinations were 'relu', 'RMSprop', 'huber loss' [54] and 'mean squared error' respectively for all three models (U-net, Res-U-net and Attention-U-net model). Where they differed was that U-net had the best combination with a 0.001 learning rate, while Res-U-net and Attention-U-net had the best combination at a 0.01 learning rate.

Table 1 Hyper-parameter variables.

Hyper-parameters	Input parameters
Learning rate	0.0001, **0.001** [1], **0.011** [1], 0.1
Activation	'sigmoid', 'tanh', **'relu'** [1]
Optimizer	'SGD', **'RMSprop'** [1], 'Adam'
Loss	'mean squared error', 'mean absolute error', **'huber loss'** [1]
Metrics	**'mean squared error'** [1], 'root mean squared error'

[1] Best performing parameter in the respective hyper-parameter category.

Further hyper-parameters like batch size, kernel size and number of epochs were decided through hit and trial or limited due to hardware constraints during this study. Batch size could not exceed past '1', as doing so resulted in a hardware bottleneck. Kernel sizes of 3 x 3 and 5 x 5 were tested, and results did not differ at this resolution; instead, the computational time in the 5 x 5 kernel was way higher than 3 x 3. The number of epochs was set to 200 for each model with early stopping enabled such that no improvement in validation loss at an iteration of 20 epochs will terminate the training. Data augmentation [55] was also applied to improve generalization capability.

 doi:10.20944/preprints202401.1879.v1

3.1.2 Performance evaluation of model losses

Loss functions play a crucial role in the training phase of the model, and understanding it gives a better insight into the black box regime of deep learning models. Since in this study three loss functions were used, namely 'mean squared error', 'mean absolute error', and 'huber loss', a training and validation loss curve gives us a better understanding of how different models behave with that loss function in this scenario.

Table 2 shows the loss functions used at the training stage of the models, where N is the number of samples, y_i is the true value, $\hat{y}_i$ is the prediction value, and δ is the huber constant, which for this study is kept at 0.5.

Table 2 Loss functions and equations

Loss function	Equation
Mean squared error	$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
Mean absolute error	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$
Huber loss	$\begin{cases} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}(y_i - \hat{y}_i)^2, & if\ \lvert y_i - \hat{y}_i\rvert \le \delta \\ \frac{1}{N}\sum_{i=1}^{N}\delta\left(\lvert y_i - \hat{y}_i\rvert - \frac{1}{2}\delta\right), & otherwise \end{cases}$

Error! Reference source not found. illustrates the training and validation loss of all three models with different loss functions incorporated: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Huber Loss, since the problem we are tackling here is much of a regression-based rather than classification-based approach. Huber loss performs better in this scenario across all the models, coming to the solution much more stable and faster than the other two regression loss functions. Since the dataset has many outliers, MSE fails prominently under these conditions, whereas MAE deals with outliers a lot better but sometimes fails to predict correctly and struggles in validation (at the early stages of training epochs and primarily in the Attention-U-net's case where the model is stagnant, i.e., not learning at all). A combination of these, the Huber loss function takes the best of MAE to deal with outliers but also incorporates proper weights for a better validation curve.

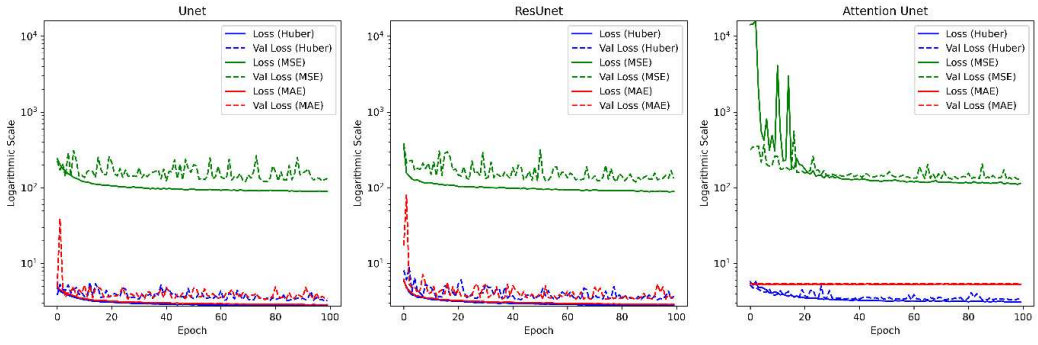

Figure 6. Loss and Validation Loss functions compared across all the models. Source: Author.

3.2 Evaluation of model predictions

A benchmark must be set for properly comparing the deep learning models tested in this study. Therefore, comparing it with the dataset currently, the only globally available 10-meter resolution fractional built-up cover is the most viable option. GHSL-S2 [34] was fetched from ESA's GHSL website, an open-access database for GHSL products. Tile R6 C26 was downloaded, corresponding

 doi:10.20944/preprints202401.1879.v1

to UTM grid zones 43R, 44R and 45R of Sentinel-2. Deep CNN architecture with transfer learning was used to generate this data. For the study area, the maximum value (i.e., fractional built-up value in the GHSL-S2 pixel) of the GHSL-S2 dataset is 0.85.

3.2.1 Qualitative Analysis

Side-by-side representations between the models are done where the predictions were visually compared in three different scenarios, (1) Urban, (2) Semi-urban, and (3) rural, each with two examples. Sentinel-2 false-color composite, LISS-4 (5-meter spatial resolution) dataset pan-sharpened with Cartosat-3 panchromatic band (0.28-meter spatial resolution) in a false-color composite and reference built-up labels were visualized together for qualitative analysis. As seen in **Error! Reference source not found.**, GHSL-S2 performed relatively poorly in all scenarios and made inaccurate predictions throughout each set, especially in urban areas. In rural settings with a minimal dataset of built-up, GHSL-S2 fails to show the built-up properly.

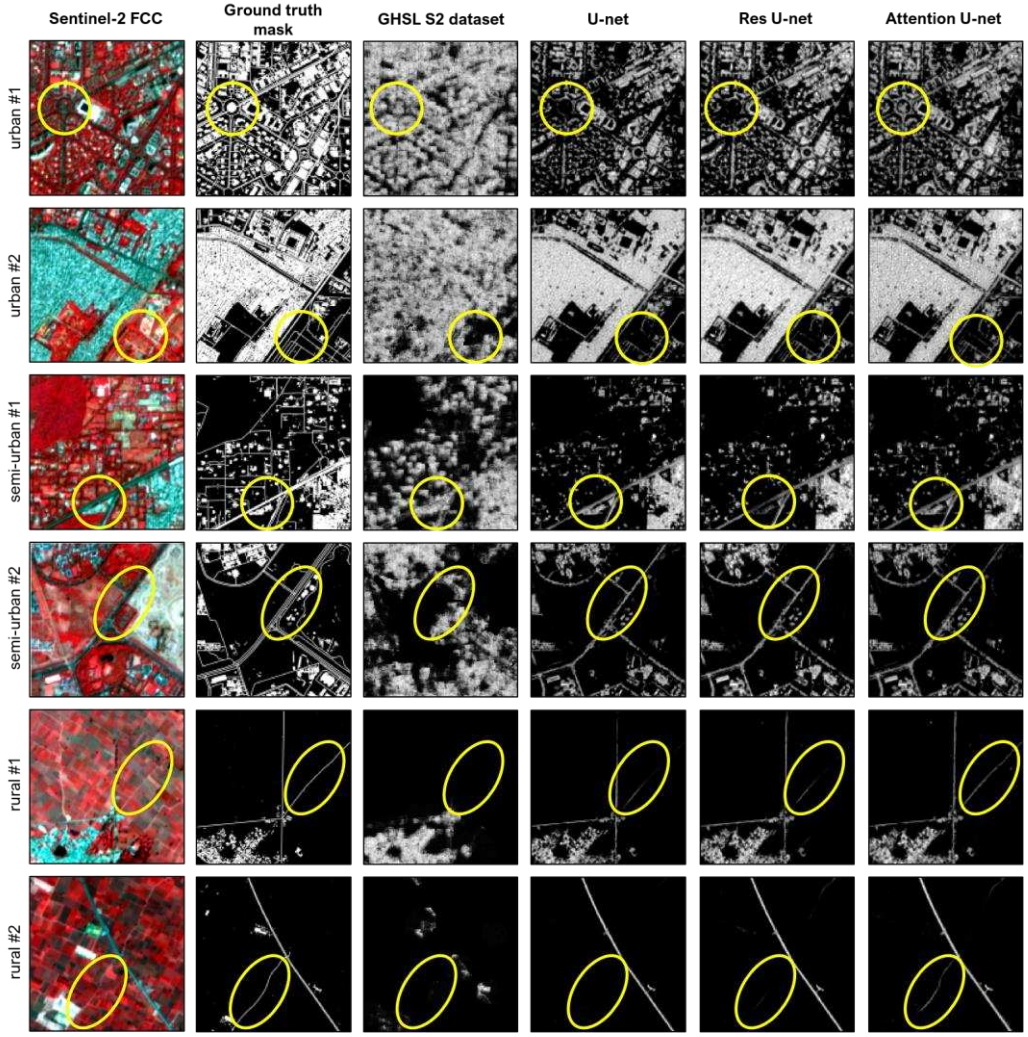

Figure 7. left to right) Sentinel-2 FCC imagery, LISS-4 FCC pan-sharpened with Cartosat-3 PAN, Reference ground truth mask, GHSL-S2 dataset, U-net prediction, Res-U-net prediction, & Attention-U-net prediction (black to white = 0% to 100% built-up; applies to all the fractional built-up and yellow circles to denote point of interests). Source: Author.

U-net performed better in urban settings but not in semi-urban and rural settings with missing roads. It can predict the roundabout in urban settings and the roads having vegetation around it, but it surpasses GHSL-S2 in accurately predicting dense urban settlements and somewhat scarce built-up. While it may be suitable for specific applications where detailed information about urban

 doi:10.20944/preprints202401.1879.v1

development is abundant and precise (such as densely populated cities), its limitations make it less reliable when dealing with low-built-up density areas. Res-U-net performed similarly to U-net, predicting densely packed built-up or dispersed with slightly better prediction ability, especially in urban and scarce rural settings. As seen in rural settings 1 and 2, the roads are slightly more accurately predicted compared to U-net, and the overall prediction of built-up is much higher (whiter = higher built-up fraction) in densely built-up areas in both urban and semi-urban settings.

Res-U-net offered comparable accuracy while providing additional benefits due to its residual connections that helped alleviate vanishing gradient problems during training. Despite these advantages over basic U-net architecture, there are still limitations regarding accurate predictions of missing or sparse fractional built-up regions. The best was Attention-U-net, with better generalization capability in all the scenarios. It is the only one to consistently predict roads much better than other models used in the study. An evident example of this can be seen in **Error! Reference source not found.**, where it is the only one able to predict the roundabout. Further, in a semi-urban setting, it can predict secluded built-up areas and small roads like U-net but with slightly better generalization. It has a much better road prediction in rural settings than the other two models.

Compared to Res-U-net, which has given more confidence to built-up in dense built-up areas (i.e., higher values and whiter pixels can be seen in dense regions of built-up), Attention-U-net goes for a more generalized and spread-out values of confidence or probability of built-up even in dense built-up regions with better edge detection capability due to its attention mapping capability [52]. This highlights its ability to understand the spread of built-up across the whole image and not only dense built-up regions (higher dataset/density of pixels) as Res-U-net. Further, this shows the superiority of attention mechanisms and attention mapping in datasets like these, where the distribution of information can be uneven.

3.2.1 Quantitative Analysis

When evaluating them based on R^2 and RMSE scores against the generated reference built-up labels dataset, GHSL-S2 scored the lowest R^2 score of 0.387 and a very high RMSE of 11.949, which correlates to the inaccurate predictions seen above by GHSL-S2. U-net, Res-U-net, and Attention-U-net scored 0.612, 0.623, and 0.631 R2 scores, respectively, while 9.913, 8.991, and 9.611 RMSE, respectively, as shown in Table 3. Overall, Attention-U-net has a slightly better fit (less variance) than other models, with 2nd best prediction regarding the RMSE score resonating with the visual comparison shown in **Error! Reference source not found.**.

Table 3 R^2 and RMSE of the models showing the best in each metric highlighted in bold. Overall, the best model is Attention-U-net. Followed by Res-U-net, U-net, and at the end is the GHSL-S2 dataset with its deep CNN model.

Model	R^2 score	RMSE
GHSL-S2 (deep CNN)	0.387	11.949
U-net	0.612	9.913
Res-U-net	0.623	**8.991**
Attention-U-net	**0.631**	9.611

3.3. Qualitative Assessment of the transferability performance

We tested the transfer ability of our models in different areas far away from the original study area, giving us a better insight into the model's generalization capability. The best-performing model, Attention-U-net, was qualitatively (visually) compared with the GHSL-S2 datasets of regions not under the study area. This experimental test lays out fundamental insights on the capability of Attention-U-net to train on a handful of data to predict from a spatially distinct Sentinel-2 dataset.

For this, five cities were meticulously chosen (three Indian and two non-Indian) to test out the model on the following criteria: a) distinguishing between elements with similar spectral signatures

 doi:10.20944/preprints202401.1879.v1

as that of built-up, i.e., barren land, shoreland, sand, etc., b) can model predict both organic and structures morphologies across urban and rural areas.

Error! Reference source not found. shows the prediction of Attention-U-net compared to the GHSL-S2 dataset in five different cities. Amsterdam shows the ability of the attention model to predict a structured morphology even though trained with an organic morphology dataset. Chennai and Mumbai both challenge the model with the large water body with clustered morphology type where, in Chennai's case, GHSL-S2 performs better than Attention-U-net, but the inverse can be said in Mumbai's case. Dubai and Mussoorie provide a geographical challenge to the Attention-U-net, with desert and hilly terrain interrupting the prediction of built-up due to similar spectral profiles, respectively. False classification can be seen prominently by Attention-U-net in both cases where it has taken sand and some parts of terrain under built-up. This can be improved through more diverse training of Attention-U-net with distinct demographics and morphologies for better prediction.

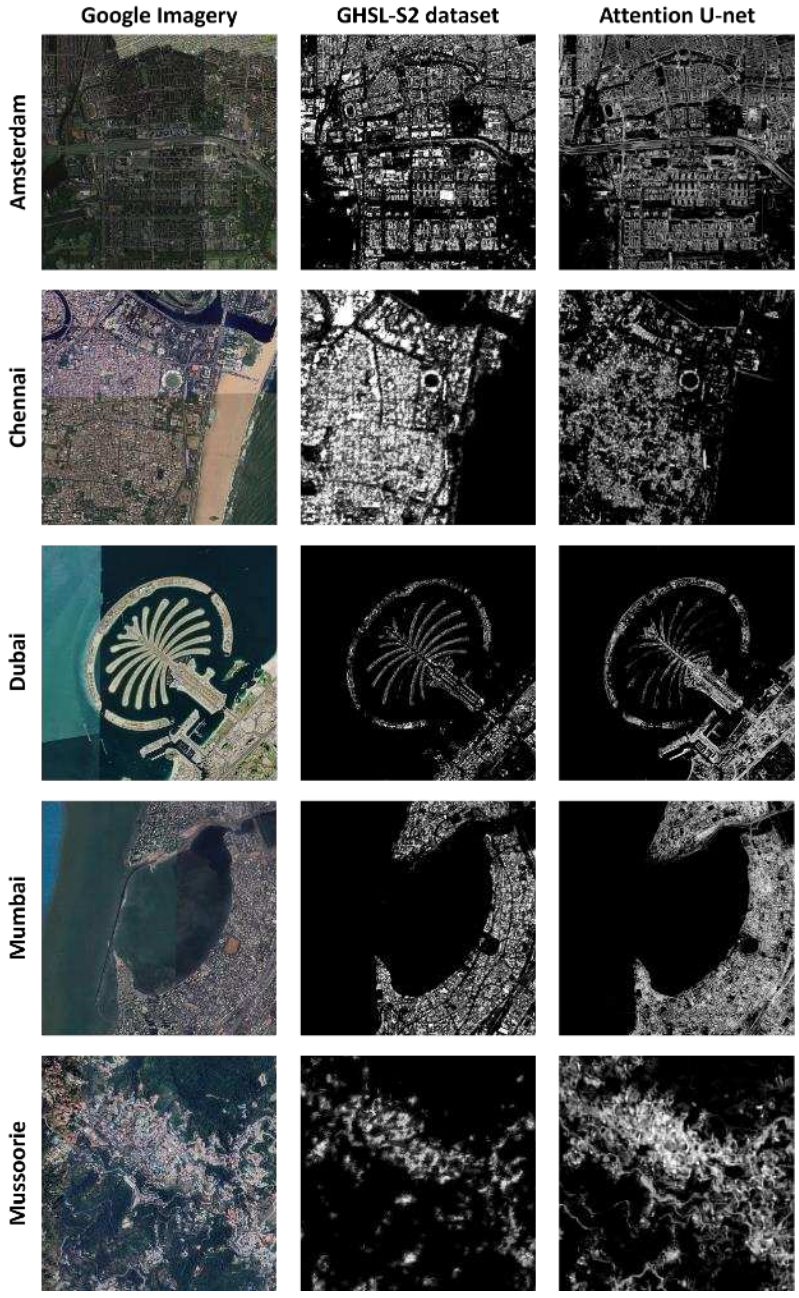

Figure 8. Generalization experimental test of Attention-U-net in different cities compared to the GHSL-S2 dataset testing different morphological patterns and geographic characteristics.

 doi:10.20944/preprints202401.1879.v1

4. Discussion

The study evaluated several deep learning models for their performance in predicting fractional built-up areas across different scenarios. The models included GHSL-S2's deep CNN, U-net, Res-U-net, and Attention-U-net. Each model was tested in urban, semi-urban, and rural settings to assess its ability to predict fractional built-up areas accurately.

GHSL-S2 performed poorly in all given scenarios when predicting fractional built-up areas. It struggled to identify and delineate these areas consistently and accurately across different settings. Although the GHSL-S2 dataset used for this study was for the period of 2018 and other models were predicting the 2021-22 period, the temporal gap between them did not drastically affect both urban and rural areas under the study region selected (see **Error! Reference source not found.**). Therefore, based on this study's findings, GHSL-S2's deep CNN is less effective than other models for predicting fractional built-up areas.

U-net performed better in urban settings than in semi-urban and rural areas with sparse or missing fractional built-up data. However, like GHSL-S2, it still fell short in accurately predicting these areas consistently across all scenarios. While it may be suitable for certain applications where detailed information about urban development is abundant and precise (such as densely populated cities), its limitations make it less reliable when dealing with incomplete or scarce data regarding fractional built-up regions.

Res-U-net exhibited similar performance to the standard U-net architecture but with slightly improved prediction ability, particularly in urban and sparsely developed rural settings. It offered comparable accuracy while providing additional benefits due to its residual connections that helped alleviate vanishing gradient problems during training. Despite these advantages over basic U-net architecture, there are still limitations regarding accurate predictions of missing or sparse fractional built-up regions.

Attention-U-net demonstrated the best overall performance among all the models studied by exhibiting superior generalization capability across various scenarios when predicting fractional built-up areas. It consistently outperformed other models in accurately identifying and delineating these regions more effectively than any other model used in this study. One of the critical strengths of Attention-U-net is its ability to utilize the wasted potential of skip connections in the standard U-net architecture. By incorporating attention mechanisms into these skip connections, the model can effectively produce responses at each pixel by weighting features from the preceding layer. This attention-based approach allows the model to handle minimal built-up elements with great accuracy, making it highly suitable for tasks involving fine-grained details in urban areas [26].

When examining statistical metrics, Attention-U-net outperforms Res-U-net and U-net in terms of R^2 score and RMSE. While the improvement in R^2 score is relatively small as seen in Table, it still demonstrates that the added complexity and computational time required for Attention-U-net is worthwhile, as it leads to better overall performance.

For more evidence in support of this evaluation of the effectiveness of deep learning models for predicting fractional built-up areas: Attention-U-net successfully identified and delineated fractional built-up areas in a densely populated urban areas (see **Error! Reference source not found.**) better than U-net, which struggled to capture these regions' complexity accurately.

In a semi-urban setting example, both U-net and Res-U-net encountered difficulties in predicting fractional built-up areas where there were missing or incomplete data (i.e., small buildings and roads alongside agricultural fields or water bodies as seen in **Error! Reference source not found.** . However, Attention-U-net generated more accurate results by effectively incorporating contextual information and learning long-range dependencies, leading to better identification of these regions even with limited data. In a rural setting with sparse fractional built-up data, GHSL-S2 performed poorly due to its inability to handle such sparse information [34]. On the other hand, both U-net and Res-U-net showed some improvements compared to GHSL-S2 but failed to predict fractional built-up areas accurately. Attention-U-net demonstrated exceptional performance by identifying small, isolated, developed regions within sparsely populated areas.

In conclusion, the analysis of Attention-U-net for fractional built-up generation highlights its superiority over other architectures in the study. Its attention mechanisms, which use the skip connections in the U-net architecture, lead to more accurate predictions, particularly in scenarios with fine-grained urban features. While there is an added computational cost, the model's performance justifies the investment, making it a powerful tool for urban-rural area analysis and prediction tasks.

5. Conclusions

In conclusion, evaluating various deep learning models for predicting fractional built-up areas has revealed important insights into their performance. Among the models examined, Attention-U-net emerged as a standout performer across different built-up densities, demonstrating its versatility and effectiveness even with a limited number of input dataset. In contrast, Res-U-net, U-net, and GHSL-S2 dataset (deep CNN) exhibited varying degrees of decrease in R^2 scores, with the latter experiencing a substantial decline. Notably, all the encoder-decoder architectures distinguished between non-built-up and built-up areas, outperforming the deep CNN model used in a prior study.

Even though this study comes with its limitations for not training with a much bigger dataset or much more complex deep-learning architecture like transformers[56] due to hardware and resource limitations it shows the capability of Attention-U-net in generating fractional built-up cover at a large scale. Attention-U-net exhibited great promise, especially in scenarios with limited built-up pixel coverage largely due to its attention block mechanism helping it distinguish the background (non-built-up pixels) from the foreground (or, built-up pixels). Future research should explore its potential further by incorporating more efficient deep learning architectures and leveraging transfer learning techniques. Even though the models were trained in the context of Indian cities with limited training datasets, they can achieve good prediction in non-Indian cities compared to the GHSL-S2 dataset as evident when Attention-U-net was tested in 5 different cities, two of which were non-Indian cities.

Additionally, the study also highlights that the GHSL-S2 dataset is the only dataset available for large-scale fractional built-up dataset as of date and more improvements need to be done leveraging open-source freely available high-resolution imagery. This study is the first to innovate over the GHSL-S2 dataset and push the effectiveness and accuracy of deep-learning models used by Corbane et al. [34] to generate large-scale fractional built-up using encoder-decoder architecture models over CNNs with open-source imageries.

Ultimately, the outcomes of this research have implications for creating national and global-scale fractional built-up datasets using multi-spectral data improving over the works of GHSL-S2's work [34]. This improved dataset offers valuable insight for a grand scheme of applications like settlement patterns[57], land management[58], slum mapping[59], disaster monitoring[25], Urban Heat Island effect[60], Urban Planning[61], and many more which can directly or indirectly use such dataset.

Author Contributions: Conceptualization, C.P., P.K.G. and A.R.; methodology, A.R.; validation, A.R., P.K.G. and C.P.; formal analysis and investigation, A.R., P.K.G. and C.P.; writing—original draft preparation, A.R.; writing—review and editing, A.R., P.K.G. and C.P.; visualization, A.R.; supervision, P.K.G. and C.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code used in this research can be obtained from Zenodo DOI: https://doi.org/10.5281/zenodo.8282636.

Acknowledgments: This research work was carried out as a part of the M.Sc. dissertation by the first author (A.R.) as a part of joint education program (JEP) of Faculty ITC, University of Twente, The Netherlands, and Indian Institute of Remote Sensing (IIRS), Dehradun. The authors are grateful to the heads of IIRS and Faculty ITC, University of Twente for the necessary facilities, support, and encouragement. Thanks to USGS EROS data center for providing free Sentinel-2a MSI data, Google for providing Earth Engine platform, as well as open-source developers for building QGIS and Python, which were used to execute this study. We thank the anonymous reviewers for their valuable and helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schneider, A.; Woodcock, C.E. Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Growth in Twenty-Five Global Cities Using Remotely Sensed Data, Pattern Metrics and Census Information. *Urban Studies* **2008**, *45*, 659–692, doi:10.1177/0042098007087340.
2. Weng, Q. A Remote Sensing?GIS Evaluation of Urban Expansion and Its Impact on Surface Temperature in the Zhujiang Delta, China. *Int J Remote Sens* **2001**, *22*, 1999–2014, doi:10.1080/713860788.
3. Theethai Jacob, A.; Jayakumar, A.; Gupta, K.; Mohandas, S.; Hendry, M.A.; Smith, D.K.E.; Francis, T.; Bhati, S.; Parde, A.N.; Mohan, M.; et al. Implementation of the Urban Parameterization Scheme in the Delhi Model with an Improved Urban Morphology. *Quarterly Journal of the Royal Meteorological Society* **2023**, *149*, 40–60, doi:10.1002/qj.4382.
4. Longley, P. Global Mapping Of Human Settlement: Experiences, Datasets, and Prospects. *The Photogrammetric Record* **2010**, *25*, 205–207, doi:https://doi.org/10.1111/j.1477-9730.2010.00574_3.x.
5. United Nations Department of Economic and Social Affairs *World Urbanization Prospects: The 2018 Revision*; UN, 2019; ISBN 9789210043144.
6. Taubenböck, H.; Esch, T.; Felbier, A.; Wiesner, M.; Roth, A.; Dech, S. Monitoring Urbanization in Mega Cities from Space. *Remote Sens Environ* **2012**, *117*, 162–176, doi:https://doi.org/10.1016/j.rse.2011.09.015.
7. Wulder, M.A.; Roy, D.P.; Radeloff, V.C.; Loveland, T.R.; Anderson, M.C.; Johnson, D.M.; Healey, S.; Zhu, Z.; Scambos, T.A.; Pahlevan, N.; et al. Fifty Years of Landsat Science and Impacts. *Remote Sens Environ* **2022**, *280*, 113195, doi:https://doi.org/10.1016/j.rse.2022.113195.
8. Zhao, Q.; Yu, L.; Du, Z.; Peng, D.; Hao, P.; Zhang, Y.; Gong, P. An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends. *Remote Sens (Basel)* **2022**, *14*, doi:10.3390/rs14081863.
9. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Change Detection Techniques. *Int J Remote Sens* **2004**, *25*, 2365–2401, doi:10.1080/0143116031000139863.
10. Varma, M.K.S.; Rao, N.K.K.; Raju, K.K.; Varma, G.P.S. Pixel-Based Classification Using Support Vector Machine Classifier. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC); February 2016; pp. 51–55.
11. Aslani, M.; Seipel, S. A Fast Instance Selection Method for Support Vector Machines in Building Extraction. *Appl Soft Comput* **2020**, *97*, 106716, doi:https://doi.org/10.1016/j.asoc.2020.106716.
12. Thottolil, R.; Kumar, U. Automatic Building Footprint Extraction Using Random Forest Algorithm from High Resolution Google Earth Images: A Feature-Based Approach. In Proceedings of the 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT); July 2022; pp. 1–6.
13. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information. *ISPRS Journal of Photogrammetry and Remote Sensing* **2004**, *58*, 239–258, doi:https://doi.org/10.1016/j.isprsjprs.2003.10.002.
14. Janalipour, M.; Mohammadzadeh, A. Building Damage Detection Using Object-Based Image Analysis and ANFIS From High-Resolution Image (Case Study: BAM Earthquake, Iran). *IEEE J Sel Top Appl Earth Obs Remote Sens* **2016**, *9*, 1937–1945, doi:10.1109/JSTARS.2015.2458582.
15. Jiang, N.; Zhang, J.X.; Li, H.T.; Lin, X.G. Semi-Automatic Building Extraction from High Resolution Imagery Based on Segmentation. In Proceedings of the 2008 International Workshop on Earth Observation and Remote Sensing Applications; June 2008; pp. 1–5.
16. Schlosser, A.D.; Szabó, G.; Bertalan, L.; Varga, Z.; Enyedi, P.; Szabó, S. Building Extraction Using Orthophotos and Dense Point Cloud Derived from Visual Band Aerial Imagery Based on Machine Learning and Segmentation. *Remote Sens (Basel)* **2020**, *12*, doi:10.3390/rs12152397.
17. Sewak, M.; Sahay, S.K.; Rathore, H. Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection. In Proceedings of the 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD); June 2018; pp. 293–296.
18. Persello, C.; Stein, A. Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 2325–2329, doi:10.1109/LGRS.2017.2763738.
19. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2014**, 3431–3440.
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015* **2015**, 234–241.
21. Yuan, J. Learning Building Extraction in Aerial Scenes with Convolutional Networks. *IEEE Trans Pattern Anal Mach Intell* **2018**, *40*, 2793–2798, doi:10.1109/TPAMI.2017.2750680.
22. Wu, G.; Shao, X.; Guo, Z.; Chen, Q.; Yuan, W.; Shi, X.; Xu, Y.; Shibasaki, R. Automatic Building Segmentation of Aerial Imagery Using Multi-Constraint Fully Convolutional Networks. *Remote Sens (Basel)* **2018**, *10*, doi:10.3390/rs10030407.

23. Sherrah, J. Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery. *ArXiv* **2016**, *abs/1606.02585*.
24. Seferbekov, S.S.; Iglovikov, V.I.; Buslaev, A. V.; Shvets, A.A. Feature Pyramid Network for Multi-Class Land Segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* **2018**, 272–2723.
25. Francini, M.; Salvo, C.; Viscomi, A.; Vitale, A. A Deep Learning-Based Method for the Semi-Automatic Identification of Built-Up Areas within Risk Zones Using Aerial Imagery and Multi-Source GIS Data: An Application for Landslide Risk. *Remote Sens (Basel)* **2022**, *14*, doi:10.3390/rs14174279.
26. Wu, Y.; Zhang, R.; Zhan, Y. Attention-Based Convolutional Neural Network for the Detection of Built-Up Areas in High-Resolution SAR Images. In Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; July 2018; pp. 4495–4498.
27. Chen, Y.; Yao, S.; Hu, Z.; Huang, B.; Miao, L.; Zhang, J. Built-Up Area Extraction Combing Densely Connected Dual-Attention Network and Multiscale Context. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2023**, *16*, 5128–5143, doi:10.1109/JSTARS.2023.3281363.
28. Tan, Y.; Xiong, S.; Li, Y. Automatic Extraction of Built-Up Areas From Panchromatic and Multispectral Remote Sensing Images Using Double-Stream Deep Convolutional Neural Networks. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2018**, *11*, 3988–4004, doi:10.1109/JSTARS.2018.2871046.
29. Zou, B.; Li, W.; Zhang, L. Built-Up Area Extraction Using High-Resolution SAR Images Based on Spectral Reconfiguration. *IEEE Geoscience and Remote Sensing Letters* **2021**, *18*, 1391–1395, doi:10.1109/LGRS.2020.3000036.
30. Bordbari, R.; Maghsoudi, Y.; Salehi, M. DETECTION OF BUILT-UP AREAS USING POLARIMETRIC SYNTHETIC APERTURE RADAR DATA AND HYPERSPECTRAL IMAGE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2015**, *XL-1/W5*, 105–110, doi:10.5194/isprsarchives-XL-1-W5-105-2015.
31. Zhang, J.; Zhang, X.; Tan, X.; Yuan, X. Extraction of Urban Built-Up Area Based on Deep Learning and Multi-Sources Data Fusion—The Application of an Emerging Technology in Urban Planning. *Land (Basel)* **2022**, *11*, 1212, doi:10.3390/land11081212.
32. Phiri, D.; Simwanda, M.; Salekin, S.; Nyirenda, V.R.; Murayama, Y.; Ranagalage, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sens (Basel)* **2020**, *12*, doi:10.3390/rs12142291.
33. Radoux, J.; Chomé, G.; Jacques, D.C.; Waldner, F.; Bellemans, N.; Matton, N.; Lamarche, C.; D'Andrimont, R.; Defourny, P. Sentinel-2's Potential for Sub-Pixel Landscape Feature Detection. *Remote Sens (Basel)* **2016**, *8*, doi:10.3390/rs8060488.
34. Corbane, C.; Syrris, V.; Sabo, F.; Politis, P.; Melchiorri, M.; Pesaresi, M.; Soille, P.; Kemper, T. Convolutional Neural Networks for Global Human Settlements Mapping from Sentinel-2 Satellite Imagery. *Neural Comput Appl* **2021**, *33*, 6697–6720, doi:10.1007/s00521-020-05449-7.
35. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens Environ* **2017**, *202*, 18–27, doi:https://doi.org/10.1016/j.rse.2017.06.031.
36. Geofabrik Download Server Available online: https://download.geofabrik.de/asia/india.html (accessed on 15 October 2023).
37. INSPIRE-Thematic Working Group Buildings *D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines*; 2013;
38. Corbane, C.; Pesaresi, M.; Kemper, T.; Politis, P.; Florczyk, A.J.; Syrris, V.; Melchiorri, M.; Sabo, F.; Soille, P. Automated Global Delineation of Human Settlements from 40 Years of Landsat Satellite Data Archives. *Big Earth Data* **2019**, *3*, 140–169, doi:10.1080/20964471.2019.1625528.
39. Pesaresi, M.; Syrris, V.; Julea, A. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sens (Basel)* **2016**, *8*, doi:10.3390/rs8050399.
40. Data for Good at Meta (Previously Facebook) - Humanitarian Data Exchange Available online: https://data.humdata.org/organization/facebook (accessed on 15 May 2023).
41. Mapping the World to Help Aid Workers, with Weakly, Semi-Supervised Learning Available online: https://ai.facebook.com/blog/mapping-the-world-to-help-aid-workers-with-weakly-semi-supervised-learning/ (accessed on 15 May 2023).
42. Tiecke, T.G.; Liu, X.; Zhang, A.; Gros, A.; Li, N.; Yetman, G.; Kilic, T.; Murray, S.; Blankespoor, B.; Prydz, E.B.; et al. Mapping the World Population One Building at a Time. *Mapping the World Population One Building at a Time* **2017**, doi:10.1596/33700.
43. Microsoft Building Footprints - Bing Maps Available online: https://www.microsoft.com/en-us/maps/building-footprints (accessed on 7 December 2022).
44. Corbane, C.; Politis, P.; Kempeneers, P.; Simonetti, D.; Soille, P.; Burger, A.; Pesaresi, M.; Sabo, F.; Syrris, V.; Kemper, T. A Global Cloud Free Pixel- Based Image Composite from Sentinel-2 Data. *Data Brief* **2020**, *31*, 105737, doi:https://doi.org/10.1016/j.dib.2020.105737.

45. Buildings - OpenStreetMap Wiki Available online: https://wiki.openstreetmap.org/wiki/Buildings (accessed on 15 May 2023).
46. Highways - OpenStreetMap Wiki Available online: https://wiki.openstreetmap.org/wiki/Highways (accessed on 15 May 2023).
47. Railways - OpenStreetMap Wiki Available online: https://wiki.openstreetmap.org/wiki/Railways (accessed on 15 May 2023).
48. GDAL/OGR contributors {GDAL/OGR} Geospatial Data Abstraction Software Library Available online: https://gdal.org.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2015**, 770–778.
50. Yi, Y.; Zhang, Z.; Zhang, W.; Zhang, C.; Li, W.; Zhao, T. Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network. *Remote Sens (Basel)* **2019**, *11*, doi:10.3390/rs11151774.
51. Wang, H.; Wang, Y.; Zhang, Q.; Xiang, S.; Pan, C. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. *Remote Sens (Basel)* **2017**, *9*, doi:10.3390/rs9050446.
52. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *European Conference on Computer Vision* **2018**.
53. Brazdil, P.; van Rijn, J.N.; Soares, C.; Vanschoren, J. Metalearning for Hyperparameter Optimization. *Metalearning* **2022**.
54. Cavazza, J.; Murino, V. Active Regression with Adaptive Huber Loss. *arXiv preprint arXiv:1606.01568* **2016**.
55. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J Big Data* **2019**, *6*, 60, doi:10.1186/s40537-019-0197-0.
56. Zhou, D.; Wang, G.; He, G.; Yin, R.; Long, T.; Zhang, Z.; Chen, S.; Luo, B. A Large-Scale Mapping Scheme for Urban Building From Gaofen-2 Images Using Deep Learning and Hierarchical Approach. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2021**, *14*, 11530–11545, doi:10.1109/JSTARS.2021.3123398.
57. Ma, L.; Guo, X.; Tian, Y.; Wang, Y.; Chen, M. Micro-Study of the Evolution of Rural Settlement Patterns and Their Spatial Association with Water and Land Resources: A Case Study of Shandan County, China. **2017**, doi:10.3390/su9122277.
58. Ngo, K.D.; Lechner, A.M.; Vu, T.T. Land Cover Mapping of the Mekong Delta to Support Natural Resource Management with Multi-Temporal Sentinel-1A Synthetic Aperture Radar Imagery. **2020**, doi:10.1016/j.rsase.2019.100272.
59. Dahiya, S.; Garg, P.K.; Jat, M.K. Automated Extraction of Slum Built-up Areas from Multispectral Imageries. *Journal of the Indian Society of Remote Sensing* **2020**, *48*, 113–119, doi:10.1007/s12524-019-01066-7.
60. Zhang, Y.; Wang, Y.; Ding, N.; Yang, X. Spatial Pattern Impact of Impervious Surface Density on Urban Heat Island Effect: A Case Study in Xuzhou, China. *Land (Basel)* **2022**, *11*, doi:10.3390/land11122135.
61. Chaturvedi, V.; de Vries, W.T. Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Science* **2021**, *5*, doi:10.3390/urbansci5030068.