Assessing Human Settlement Sprawl in Mexico Via Remote Sensing and Deep Learning

Antonio Briseño (iD), Joaquín Salas (iD), *Member, IEEE*, Ranyart R. Suarez (iD), Elio Villaseñor (iD), and Danielle Wood (iD)

Abstract—Understanding human settlements' geographic location and extent can support decision-making in resource distribution, urban growth policies, and natural resource protection. This research presents an approach to assess human settlement sprawl using labeled multispectral satellite image patches and Convolutional Neural Networks (CNN). By training deep learning classifiers with a dataset of 5,359,442 records consisting of satellite images and census data from 2010, we evaluate sprawl for settlements across the country. The study focuses on major cities in Mexico, comparing ground truth results for 2015 and 2020. EfficientNet-B7 achieved the best performance among various CNN architectures evaluated with an ROC AUC of 0.970 and a PR AUC of 0.972. To assess human settlement sprawl, we introduce an information-based metric that offers advantages over entropy-based alternatives. We demonstrate its application to major cities in Mexico as examples.

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8487

Index Terms—human footprint, settlements sprawl, urban growth

I. INTRODUCTION

According to the United Nations [1], the world population is projected to reach 9.7 billion by 2050 and 10.9 billion by 2100, with varying growth rates across countries. While sub-Saharan African countries are expected to maintain population growth through 2021 and host over fifty percent of global population increase through 2050, other regions such as Eastern and South-Eastern Asia, Central and Southern Asia, Latin America and the Caribbean, as well as Europe and Northern America may reach a population peak and begin to decline by 2100 [2]. This population growth will primarily occur in urban areas [3]. Effective monitoring systems that provide real-time information are crucial to support sustainable policy development. With the increasing availability of remote sensing data and advancements in machine learning (ML) techniques, analyzing urbanization patterns can be efficiently performed.

Deep Learning [4] is a leading approach for computer vision, natural language processing, and ML tasks. Its ability

Antonio Briseño and Joaquín Salas are with the Instituto Politécnico Nacional: Querétaro, Querétaro, México (e-mail: red2013@hotmail.com and salas@ieee.org).

Ranyart R. Suarez and Elio Villaseñor are with the Instituto Nacional de Estadística y Geografía: Aguascalientes, Aguascalientes, México (e-mail: ranyart.suarez@inegi.org.mx and elio.villasenor@inegi.org.mx).

Danielle Wood is with the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA (e-mail: drwood@mit.edu).

to automatically extract features from data [5] reduces bias introduced by feature engineering, resulting in high-performing applications that can rival humans [6]. However, the returns on performance improvement diminish as model complexity increases [7]. Recent advancements in machine architectural design show promise in surpassing large convolutional neural networks (CNNs). In this research, to assess human settlement sprawl dynamics on a large scale, we extensively evaluated deep learning architectures in a classification task involving millions of records.

Furthermore, in recent years, Earth observations and remote sensing have become promising tools to measure and monitor the Sustainable Development Goals (SDG) [8], [9]. Specifically, a human settlements layer has been found helpful as input to estimate indicators associated with SDG 7 (Ensure access to affordable, reliable, sustainable, and modern energy for all) [10], SDG 11 (Make cities and human settlements inclusive, safe, resilient and sustainable) [11] and SDG 15.3 (combat desertification, restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land degradation-neutral world) [12].

This research introduces a novel approach for remote monitoring of human settlement sprawl using Landsat satellite images and a comprehensive collection of geolocated labels (see Fig. 1). Using CNNs, we establish the correlation between multispectral satellite images and reference occupancy labels to analyze the longitudinal evolution of human settlements sprawl in Mexico. The contributions of this research include

- an information-based methodology for evaluating the evolution of the human settlement sprawl in Mexico.
- code and dataset publicly available, enabling other researchers to verify the results and facilitating further study in this area.
- evaluation of the dataset with CNN-based learning machines to assess their effectiveness in detecting human settlements.

The manuscript offers valuable insights into remote monitoring techniques and CNN-based learning machines for tracking human settlement sprawl, with implications for urban planning and policy-making. The remaining sections of the paper include a literature survey, dataset and methodology description, presentation of results, detailed discussions of findings, and future research directions.

II. RELATED WORK

This section provides a literature survey on human settlement detection and sprawl dynamics assessment using remote
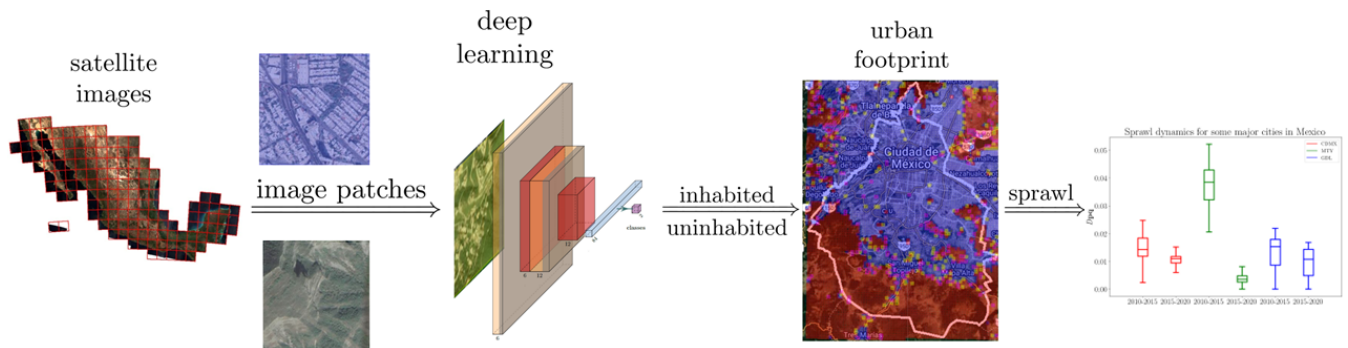
Fig. 1. A model is trained to differentiate between inhabited and uninhabited areas using satellite images and public records, enabling a temporal analysis of human sprawl.

sensing. We explore two main approaches: classical techniques and deep learning (see Table I). Our focus is specifically on assessing human settlement sprawl.

A. Human Settlements Detection

We provide a survey of literature on automatic human settlement detection using remote sensing, covering classical and deep-learning-based techniques.

Classical Approaches. Various methods have been proposed for the automated detection of human settlements using remote sensing data. Marconcini *et al.* [13] combine temporal statistics from radar imagery and spectral indices from optical imagery, employing a Support Vector Machine (SVM) classifier. A study by Hoeck and Friedrich [14] focus on small settlements and integrates four products with resolutions ranging from 50 cm to 10 m per pixel. Rudiastuti *et al.* [15] and Ayala *et al.* [16] utilize Sentinel-2 data to differentiate buildings, roads, and houses to detect human settlements. Departing from these methods, Gong *et al.* [17] propose an algorithm based on the Normalized Difference Vegetation Index (NDVI) and the Modified Normalized Difference Water Index (MNDWI) for Landsat imagery, resulting in a time-series human settlement layer spanning 40 years. Finally, Tingzon *et al.* [18] generate time series of informal settlements, employing Random Forests (RF) and SVM algorithms for pixel-wise classification.

Modern Approaches. Several studies have explored CNNs for remote sensing. Corbane *et al.* [19] and Qiu *et al.* [20] used Sentinel-2 imagery and a CNN for pixel-wise classification with the Global Human Settlement Layer as reference labels. Rapuzzi *et al.* [21] used Sentinel-1 (SAR) imagery to detect built-up areas with high and constant reflection. Qiu *et al.* [22] performed multispectral imagery classification with a ResNeXt CNN. Recent trends involve deep learning with multiple sources. For example, Wu *et al.* [23] used a semantic segmentation U-Net for built-up area detection from SAR images. Zitzlsberger *et al.* [24] processed multispectral and SAR images to monitor urban change using a CNN and an RNN. Fibæk *et al.* [25] detected human structures using a CNN with range, optical, and radar observations. Some studies explored modifications to CNN architectures. Ansari *et al.* [26] added contourlet for directional information and wavelet-based image decomposition and reconstruction.

Ghaffarian *et al.* [27] implemented attention mechanisms to improve remote sensing applications. Fan *et al.* [28] adopted multimodal remote sensing, using VHR satellite imagery and time series from population density data, applying a ResMixer, PDNet, and a transformer encoder.

Osorio *et al.* [29] utilize remote sensing techniques like lidar and photogrammetry to reveal significant ancient landscape changes in prehispanic settlements in Colombia, challenging the notion of these areas being pristine. Abate *et al.* [30] also use similar technologies for analyzing medieval settlements in Italy, focusing on producing detailed 3D and 2D data to uncover and interpret hidden sites. Aamir *et al.* [31] apply the U-net architecture, a deep learning framework, for segmenting rural settlements in Morocco, demonstrating its utility in land management. Crivellari *et al.* [32] explore using Generative Adversarial Networks (GANs) to improve satellite image resolution for detecting unplanned urban settlements in Chinese urban villages. In contrast, Alrasheedi *et al.* [33] emphasize integrating local knowledge with remote sensing to identify informal settlements accurately. These studies highlight the importance of advanced techniques in understanding various types of human settlements across different historical and geographical contexts.

B. Human Settlements Sprawl Assessment

Unmanaged Human settlements sprawl is associated with negative impacts, including increased costs for services such as water, sewage, and waste management, reduced efficiency due to strained infrastructure, higher traffic, air pollution, loss of community sense, and decreased green spaces [35]. Researchers have employed remote sensing and ML techniques to quantify these impacts. Some researchers [36]–[38] employ land-use change alone to indicate settlement sprawl, while others [39] focus on changes occurring radially from the settlement center. In addition to remote sensing, some researchers [40]–[42] have incorporated census information [43]–[45]. However, defining human settlements sprawl in terms of a single indicator may not capture the full complexity of the phenomenon. Some studies have proposed using multiple indicators and multivariate analysis to address this. For example, Gielen *et al.* [35] introduced an independent component analysis-based index constructed from 12 indicators, while Chettry *et al.* [46] employed 11 descriptors and multivariate

TABLE I

PERFORMANCE OF ML MODELS IN DETECTING HUMAN SETTLEMENT SPRAWL: FCNN (FULLY CONVOLUTIONAL NEURAL NETWORK), RF (RANDOM FOREST), NB (NAIVE BAYES), XGB (EXTREME GRADIENT BOOSTING), NN (NEAREST NEIGHBOR), SVM (SUPPORT VECTOR MACHINES)

Paper	Elements detected	Pixel resolution	Algorithms	Accuracy
[22]	settlements	80 m	ResNeXt Attention	91.0%
[13]	settlements, water bodies	10 m	SVM	89.3%
[19]	settlements	10 m	FCNN	87.5%
[34]	settlements	0.5 m	U-Net	95.0%
[20]	settlements	20 m	FCNN	90.1%
[15]	roads, settlements, water bodies	10 m	RF, NB, NN, XGB, SVM	80%, 29.0%, 48.6%, 80.0%, 71.4%
[16]	settlements, roads	2.5 m	U-Net + ResNet-34	54.4%

analysis. Other studies have used standardized G_i statistics and urban expansion intensity index based on chi-square analysis [47] to measure settlement sprawl. Bhattacharjee *et al.* [48] study on identifying potential urban and rural growth centers, illustrates the importance of employing diverse information sources like hierarchical settlements, remote sensing, and GIS for organized development of human settlements. The study underscores the necessity of considering multiple factors such as groundwater potential and land use to prevent urban sprawl, promote sustainable development, and address economic imbalances and rural migration.

Entropy is valuable for quantifying sprawl, capturing the complexity and randomness of urban development patterns. Shannon entropy has been employed by several authors [49]–[51]. Additionally, Padmanaban *et al.* [52] have used Renyi entropy, a generalization of Shannon entropy. While entropy provides insights for static analysis, studying the dynamics of urban sprawl requires divergence measurements like the Jensen-Shannon divergence. With appealing properties such as symmetry and the metric nature of the square root of twice the Jensen-Shannon divergence, this measure is suitable for comparing changes in urban development patterns over time [53].

The literature on human settlement detection and sprawl dynamics using remote sensing extensively covers various detection methods and sprawl assessments but needs a focus on using information-based metrics for comparative analysis of sprawl over time. This gap highlights the need for integrating metrics like Shannon and Jensen-Shannon entropy in a cohesive framework to quantify and compare the dynamics of urban sprawl across different periods and regions more effectively.

III. MATERIALS AND METHODS

Fig. 1 illustrates the methodological process of using satellite imagery to identify urban sprawl. First a CNN deep learning model is trained using administrative data and labeled satellite imagery. To estimate urban sprawl in new areas, unlabeled satellite images are divided into smaller image patches. These patches are processed through the CNN deep learning model that classifies them into inhabited or uninhabited areas. The output is an estimated urban footprint, which discriminates the spread of inhabited areas within a region (in the illustration around Mexico City). The final part of the figure shows a boxplot representing the sprawl dynamics for major cities in Mexico over time, indicating the rate of urban expansion. This section discusses the satellite imagery input, public dataset labels used as output, and benchmarked CNNs for identifying human settlements sprawl.

A. Satellite Imagery

This research utilizes the geomedian [54], an image incorporating statistical analysis to determine the intensities $\boldsymbol{\mu}$ for each pixel's location. The geomedian is obtained by analyzing a sequence of observations, specifically n multispectral samples denoted as $\mathbf{x}_i \in \mathbb{R}^p$, where $i = 1, \ldots, n$, and corresponding to the p-bands of a particular location. The geomedian is defined as the point $\mathbf{x}$ that minimizes the expression

$$\boldsymbol{\mu} = \arg \min_{\mathbf{x}} \sum_{i=1}^{n} ||\mathbf{x} - \mathbf{x}_i||. \qquad (1)$$

The geomedian image employed in this research is built from multiple Landsat scenes captured over the same geographic area, covering one year of observations. It results in a cloud-free satellite imagery dataset comprising pixel time series. This dataset ensures the preservation of geometrical and multispectral relationships [55].

B. Inhabited Areas in Mexico

Since 1950, Mexico has conducted population and housing censuses, defining housing as the "space delimited by walls and ceiling made by any material used to live" [56], even if not intended for lodging. The Mexican census classifies property types for each residential block, including particular housing, collective housing, particular housing with business premises, active business premises, non-active business premises, public plazas, schools, vacant lots, or buildings. In our dataset, we summarize these registrations as the inhabited class. We utilized the national grid used by the Mexican Institute of Statistics and Geography (INEGI) [57], which divides the territory into regular-sized tiles with unique identifiers. We obtained the inhabited and uninhabited classes by intersecting this grid with the shapefile published by INEGI as open data from the 2010 national census. The inhabited class consists of tiles derived from the 2010 census, while the uninhabited class comprises the remaining regions, such as space-free buildings, water bodies, scrub, jungles, and forestry. The resulting inhabited and uninhabited sets contain 333,416 and 5,026,026 images, respectively. Note that the national grid refers to a standardized geographic coordinate system used to divide a region into smaller sections for mapping purposes.
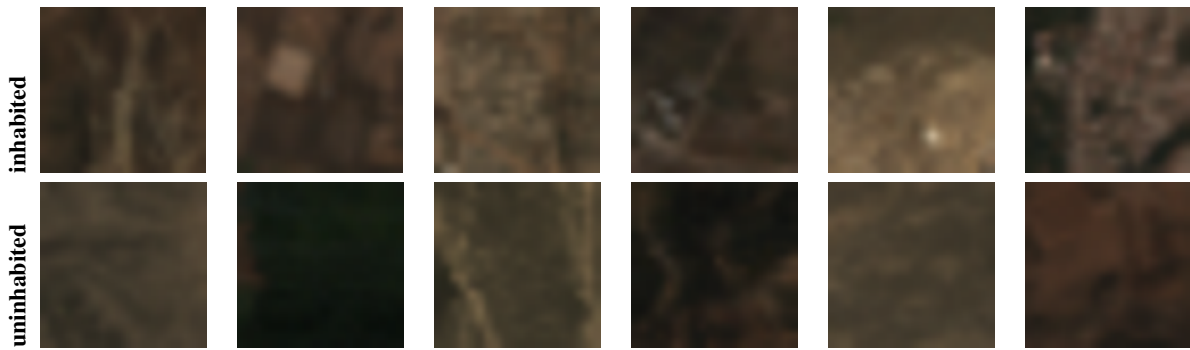
Fig. 2. Geomedian satellite imagery samples for inhabited (a) and uninhabited (b) classes were created using RGB patches from six bands. HSV representation was used, and intensity values were adjusted through histogram equalization.

C. Architectures Employed

CNNs are highly effective in computer vision tasks like image classification and object detection. This study employs CNN architectures to detect human settlements in satellite images. We start with ConvMixers [58], which have fewer parameters but still perform well. Next, we explore larger and more complex models: EfficientNet-B3 (10,787,695 parameters), EfficientNet-B7 (64,104,543 parameters), ResNet-200 (91,338,112 parameters), and an U-Net framework contructed on an EfficientNet backbone for semantic segmentation. The objective is to assess the effectiveness of different CNN architectures in detecting human settlements and identify the optimal approach for this task given the proposed dataset.

ConvMixer [58] combines the strengths of vision Transformers [59] and Multi-Linear Perceptron (MLP) Mixers [60]. It operates on image patches, maintains resolution, and extracts channel-wise and spatial-wise features. However, ConvMixer achieves these operations using convolutions, leading to significant architectural optimization. It consists of a patch-embedding layer, depth-wise and point-wise convolution blocks with non-linear activation and batch normalization layers. The network concludes with global average pooling and a fully connected layer with softmax output.

ResNet was the pioneering architecture that demonstrated the effectiveness of skip connections in improving network performance [61]. Each residual function in ResNet consists of three convolutional blocks: two with a 1×1 kernel size to adjust image dimensions and one with a 3×3 kernel size serving as an information bottleneck. Deeper ResNet models have shown superior performance to shallow ones, such as ResNet-101, outperforming ResNet-32 on the ImageNet dataset. However, there is a depth limit, as demonstrated by ResNet-1202, which exhibits overfitting [61]. The ResNet family has been further enhanced through scaling strategies and training methods, achieving faster speeds ($1.7\times$ - $2.7\times$) than the EfficientNet family while maintaining accuracy [62].

EfficientNet [63] is an innovative architecture that achieves state-of-the-art performance in various computer vision tasks. Unlike traditional ConvNets, EfficientNet strikes a balance in scaling network dimensions (width, depth, and resolution) to achieve optimal trade-offs between accuracy and efficiency. The width parameter increases the channels in the CNN,

making the network wider. The depth parameter determines the number of layers in the model, enabling it to capture more features, although deep architectures may face the vanishing gradient problem during training. Lastly, resolution refers to the input image's dimensions (width $\times$ height). Higher-resolution images enhance model performance but also require more resources.

U-Net architectures [64] enable multiple simultaneous predictions. They consist of contracting and expanding layers, with interconnections at corresponding feature resolution levels. U-Net has demonstrated success in semantic image segmentation and has been enhanced with advanced backbone networks like EfficientNet. This study uses the Eff-UNet architecture, which combines U-Net with the EfficientNet backbone [65].

D. Sprawl of Human Settlements

Entropy in information theory quantifies the average level of uncertainty. In the context of human settlements sprawl, entropy can serve as a measure of their occurrence in random locations. Lower entropy values indicate a more concentrated distribution, while higher values signify dispersion. Shannon entropy can be defined as

$$H = -\sum_{c=1}^{C} p_c \log p_c, \qquad (2)$$

where $\mathbf{p} = [p_1, \ldots, p_C]$ represents the mass probability function with individual probabilities p_c, and C represents the number of different classes. If the true probability mass is inaccessible, we can approximate the values using class frequencies, where $p_c \approx n_c/N$, with n_c denoting the occurrence of class c and N representing the total number of possible locations. In some cases, the normalized entropy $H_n = H/\log C$ may be preferred, as it provides a value between zero and one.

When evaluating human settlement dynamics, describing the probability distributions $\mathbf{p}$ and $\mathbf{q}$ for two different time instances in the same locality is important. The Kullback-Leibler (KL) divergence can be used to measure the difference between these distributions, defined as

$$D_{\mathrm{KL}} \left(\mathbf{p} \parallel \mathbf{q} \right) = \sum_{c=1}^{C} p_c \log \left(\frac{p_c}{q_c} \right). \qquad (3)$$

The KL divergence is not symmetric, *i.e.*, $D_{\mathrm{KL}}\left(\mathbf{p} \parallel \mathbf{q}\right) \neq D_{\mathrm{KL}}\left(\mathbf{q} \parallel \mathbf{p}\right)$, and does not obey the triangle inequality [66]. However, the Jensen-Shannon divergence (JSD) is considered for backward comparison. It is defined as

$$\mathrm{JSD}\left(\mathbf{p} \parallel \mathbf{q}\right) = \frac{1}{2}D_{\mathrm{KL}}\left(\mathbf{p} \parallel \mathbf{m}\right) + \frac{1}{2}D_{\mathrm{KL}}\left(\mathbf{q} \parallel \mathbf{m}\right), \quad (4)$$

where $\mathbf{m} = \frac{1}{2}\left(\mathbf{p} + \mathbf{q}\right)$. Endres *et al.* [53] studied the square of the Jensen-Shannon divergence, denoted as $D_{\mathbf{pq}}^2$, which is twice the JSD, *i.e.*,

$$D_{\mathbf{pq}}^2 = 2\mathrm{JSD}\left(\mathbf{p} \parallel \mathbf{q}\right). \quad (5)$$

They showed that $D_{\mathbf{pq}}$ is a metric. In the context of human settlements sprawl, the distributions spread across images, and thus, $D_{\mathbf{pq}}^2$ provides a descriptor of distribution differences interpreted in that bi-dimensional space.

IV. RESULTS

This section presents the performance of the various CNNs analyzed in detecting human settlements in satellite images and the assessment of sprawl for major cities in Mexico.

A. Dataset and Computing Resources

We compiled a comprehensive dataset containing 5,359,442 Landsat satellite images for this study. Each image covers an area of 600×600 m^2 at a resolution of 20×20 pixels. A random sample of examples is illustrated in Fig. 2. The images were obtained using the geomedian [55] from Landsat 5 and 7 images collected for 2010. The dataset comprises six spectral bands: red (0.63-0.69 μm), green (0.52-0.60 μm), blue (0.45-0.52 μm), near-infrared (0.77-0.90 μm), short-wave infrared one (1.55-1.75 μm), and short-wave infrared two (2.08-2.35 μm).

Of the 5,359,442 images, 5,026,026 were labeled as *uninhabited*, and the remaining 333,416 images were labeled as *inhabited*. To address the class imbalance, we randomly sampled images from the uninhabited class to match the number of images in the inhabited class. As a result, each training iteration consisted of 666,832 images, with an equal number of images (333,416) from both categories. The dataset was split into three sets: 333,416 images for training (50%), 166,708 records for validation (25%), and 166,708 records for testing (25%). This division ensured the model was trained on a large and diverse set of images while avoiding overfitting.

This study used two high-performance computing systems to train the models. The first system has an x86_64 architecture, a 24-core Intel Xeon E-2650 v4 CPU, 503 GB DDR4 RAM, and four 12 GB Titan X (Pascal) GPUs. We employed this system to train ConvMixer, and EfficientNet-B3. The second system has an x86_64 architecture, a 24-core Intel Xeon E3-1220 E5-2450 CPU, 334 GB DDR4 RAM, and eight 16 GB TPUs v3. We used this system to train EfficientNet-B7, ResNet-200, and Eff-UNet.

B. Architectures Fine-Tuning

ConvMixer. To implement ConvMixer, we utilized Keras and followed the description provided by Trockman and Zico [58]. To explore the space of hyperparameters, we trained 145 models with distinct combinations. We systematically searched over a range of values for each parameter to determine the optimal hyperparameters. Specifically, we tested different learning rates, including $5 \times 10^{-3}, 1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}$, and 1×10^{-5}. We also varied the batch size, exploring 64, 128, 256, and 512. In addition, we tried different kernel sizes, including 2×2, 5×5 and 10×10, as well as different patch sizes, employing 2×2, 5×5, and 10×10. Finally, we tested 10, 12, 15, and 20 blocks for the filter's depth. Our results indicate that starting at 15 blocks, the model increased its performance. Ultimately, the best-performing hyperparameters were identified as a learning rate of 5×10^{-4} and a batch size of 128.

Other hyperparameters include twenty depth-wise and point-wise convolutions blocks, which involve a GELU activation function, batch normalization, and a skip connection that goes before applying the depth-wise convolution and finishes after this operation. Regarding the patch-embedding process, a kernel size of 5×5 and a patch size of 2×2 were used. The resultant features were flattened in the last layers by applying fully connected layers, with a softmax function on the last one to obtain a probability distribution between the two classes. The loss function employed was binary cross-entropy. The Adam optimizer was utilized because it dynamically varies the learning rate as training progresses. During the training process we monitored the accuracy. This architecture had 5,108 parameters and took four hours to train. Initially, it was supposed to train for 100 epochs, but we applied an early stopping callback that monitored the validation accuracy to improve at least 1% every 15 epochs. As a result, training stopped on epoch 48. Both training and validation accuracy curves follow a steady upward trend. The accuracy validation curve did not present abrupt jumps and oscillated between 86% and 88%.

EfficientNet-B3. Another architecture analyzed in this work was EfficientNet-B3, which involved implementing transfer-learning with ImageNet [67] pre-trained weights. Since EfficientNet requires images of 300×300 pixels or larger, we up-sampled the 20×20 pixels images using inter-cubic interpolation. We also modified the architecture because EfficientNet layers with ImageNet weights cannot receive six-channel images. Therefore, we duplicated EfficientNet without ImageNet weights and changed the first four layers to process the six channels from the images. Additionally, we remove the top layer, which consists of a Dense layer that converts the 1,280 features into the 1,000 ImageNet classes, and replace it with a Global Average Pooling layer.

Continuing with the last layers, we added a Dropout layer with a rate of 0.2 to reduce overfitting, followed by a Dense layer with a softmax activation function to predict the probabilities between the two classes. The hyper-parameters used in this experiment were similar to ConvMixer. The learning rate of 1×10^{-6}, and a batch size of 64. We utilized the

(a) ROC Curves

(b) PR Curves
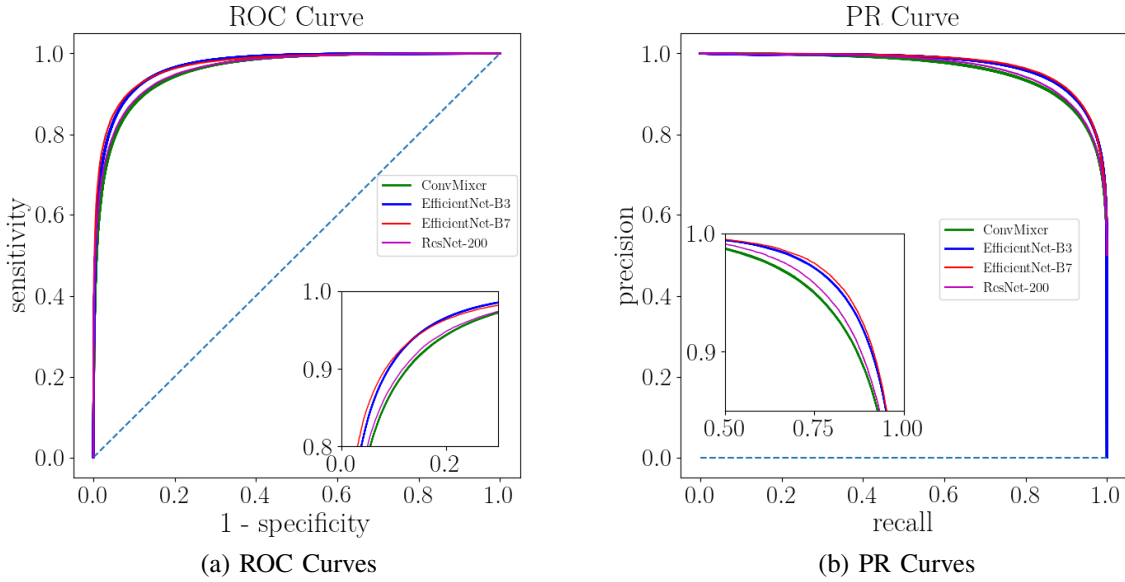
Measure		ConvMixer	EfficientNet-B3	EfficientNet-B7	RestNet-200
ROC AUC	μ	0.957	0.968	0.970	0.960
	σ	2.0×10^{-4}	1.5×10^{-4}	n/a	n/a
PR AUC	μ	0.958	0.969	0.971	0.962
	σ	1.8×10^{-4}	1.0×10^{-4}	n/a	n/a

(c)

Fig. 3. CNN statistical performance on the task of detecting human settlements. Performance curves of 14 tests were evaluated using a dataset of 666,832 images from both uninhabited and inhabited samples. ConvMixer, and EfficientNet-B3 models were trained with different data splits. A t-test reveals that the performance is statistically different with a p-value < 0.01. EfficientNet-B7 and ResNet-200 were only run twice due to their heavy computational burden, and we found no benefit in repeating the training process given the observed result.

Adam optimizer with binary cross-entropy as the loss function. The EfficientNet-B3 with modified architecture had a total of 10,787,695 parameters. It took 33 epochs, or roughly 108 hours, to finish the training, considering that early stopping concluded it. The validation curve initially presented higher accuracy than training but converged to an accuracy of 90% as the epochs progressed.

EfficientNet-B7. This model used pre-trained weights from ImageNet for its 64,104,543 parameters that are distributed across 813 layers. To achieve better results, we rescaled the images to a size of 600×600 pixels. Following a similar approach to EfficientNet-B3, we modified the architecture by removing the final layers and replacing them with a Global Average Polling layer. We also added a Dropout layer to prevent overfitting and a Dense layer with a softmax function for the final classification step.

To facilitate comparison between different CNN architectures, we adopted a standardized approach. We used the Adam optimizer with a learning rate of 1×10^{-6} and a batch size of 64. Binary cross-entropy served as our loss function. We trained the model for 70 epochs, implementing early stopping and saving the model after each epoch. After 29 training epochs, we obtained the best model, which achieved a validation accuracy of approximately 91.1%.

ResNet-200. This model has 91,338,112 trainable parameters and requires an input image size of 256×256. We modified the architecture to ensure consistency and enable performance comparison across models. For instance, we adjusted the first three layers to accommodate the six channels of our dataset and utilized pre-trained ImageNet weights. We also replaced the fully connected layers with a Global Average Pooling layer and incorporated a softmax activation function. To train the model, we reused the hyperparameters from previous experiments, including a batch size of 64, a learning rate of 1×10^{-6}, and binary cross-entropy as the loss function. After 70 epochs, we achieved the best weights at the 26th epoch, with a validation accuracy of 89.9%.

U-Net. For this model, we constructed mosaics by aggregating multispectral patches of size 20×20 pixels and their corresponding labels in a semantic segmentation manner. This strategy resulted in a multispectral mosaic of 520×520 pixels, containing the labels *inhabited* and *uninhabited* in mosaic patches of size 26×26 pixels. To align with the Eff-UNet architecture, we resized the input to 512×512 pixels and evaluated the performance on corresponding 25×25 pixel labels. We generated 2,519 images from the area where the initial 5,359,442 image patches were extracted.

We used the EfficientNet-B7 backbone, the best-performing

classifier in our experiments, to implement the U-Net architecture. As before, we loaded the pre-trained ImageNet weights to the EfficientNet-B7 architecture. However, we made some modifications to the backbone, including changing the input size to accept the six bands and adapting the decoder to generate a 26×26 pixel output, which we passed through a softmax activation function in the final layer. For the hyperparameters, we set the loss function to binary cross-entropy with a learning rate of 10^{-5} and a batch size of 16. To account for class imbalance, we replaced accuracy with the intersection over union (IoU) metric as a better performance indicator. We trained the model for 250 epochs but employed early stopping to save the best-performing weights obtained at epoch 122. The final IoU performance of the architecture was 83%.

C. Human Settlement Detection Results

The results demonstrate that the CNN architectures accurately classify whether an image belongs to an uninhabited or inhabited area. EfficientNet-B7 achieved the best performance with a ROC AUC of 0.970 and a PR AUC of 0.971, followed by EfficientNet-B3 with a ROC AUC of 0.968 and a PR AUC of 0.969 . By comparison, ResNet-200 was above ConvMixer with a ROC AUC of 0.960 and a PR AUC of 0.962 against a ROC of 0.957 and a PR AUC of 0.958.

To assess the classifiers' uncertainty , we trained each architecture 14 times using a subset of inhabited images sampled randomly without replacement while keeping the set of uninhabited images constant. This procedure resulted in a dataset of 666,832 records for each iteration. During each iteration, ROC and PR curves were generated for each architecture, and their performance was evaluated (see Fig. 3). EfficientNet-B7 outperformed EfficientNet-B3, which surpassed ResNet-200, while ResNet-200 outperformed ConvMixer. A t-test, with a p-value < 0.01, confirmed that models' performances differ statistically.

We conducted an experiment to demonstrate the effectiveness of the EfficientNet-B7 classifier, the best-performing classifier (see Fig. 4). We evaluated every 600×600 m^2 of the national grid that contained at least one pixel inside the territorial limits of Mexico City, Guadalajara, and Monterrey, the three largest cities in Mexico. Correspondingly, we predicted their land use for the years 2010, 2015, and 2020. While the year 2010 contained observations employed during the training process, the datasets for 2015 and 2020 contained unseen images. We evaluated the model's performance against publicly available census information to obtain ground truth for these years. The results are summarized in Fig. 5. We achieved ROC AUC above 0.984 and PR AUC above 0.975 for these cities, indicating better performance than for the country of Mexico.

We characterized the human settlement footprint for each city over the observed period (see Fig. 5). However, establishing a precise threshold for the ML inference process can lead to false positives and false negatives. Therefore, a trade-off is required to identify the optimal threshold and associated costs. Different cutoff values can be applied to evaluate

the classifier's overall performance, resulting in a different proportion of occupied settlements. The resulting distributions can be represented as boxplots, where the interquartile range (IQR) extremes Q_1 and Q_3 correspond to the 25th and 75th percentiles, respectively.

The analysis of the results revealed interesting insights about each city's settlement patterns. For instance, the IQR extremes for Mexico City indicate a proportion of inhabited areas between 0.7 to 0.75 in 2010. The median value slightly decreased by 2015 but returned to a similar proportion in 2020. In contrast, Monterrey started in 2010 with a ratio between 0.66 and 0.68 for the extremes at Q_1 and Q_3, and it increased in 2015, being 0.69 in Q_1 and 0.72 in Q_3. Then, in 2020, the extremes defined by Q_1 and Q_3 corresponded to 0.69 and 0.73. Finally, Guadalajara already had a high proportion of inhabited areas in 2010, with the IQR ranging from 0.964 to 0.969. However, by 2015, it decreased, changing the IQR from 0.963 to 0.966. In 2020, the IQR was from 0.961 to 0.968, indicating that settlements continued to occupy most of the area.

D. Human Settlement Sprawl Results

Our formulation enables the analysis of the dynamics of urban sprawl in urban settlements. To estimate this, we computed $D_{\mathbf{pq}}$ for Mexico City, Monterrey, and Guadalajara. To calculate it, we randomly sampled the distribution corresponding to the respective proportion of area with settlements for one period, compared it with another random sample from the second period, repeated this process 1,000 times, and eventually estimated the resulting distributions (see Fig. 5(b)). The results revealed that for the three cities, $D_{\mathbf{pq}}$ grew more in the period 2010-2015 relative to the period 2015-2020. The $D_{\mathbf{pq}}$ distribution IQR for Mexico City was between 0.012 and 0.0186 from 2010 to 2015 and 0.009 to 0.0117 from 2015 to 2020. Monterrey showed an IQR 2010-2015 with $Q_1 = 0.0321$ and a $Q_3 = 0.0429$. On the other hand, the IQR for 2015 to 2020 was bounded by $Q_1 = 0.0027$ and $Q_3 = 0.0047$. Guadalajara, the city with the most significant proportion of area with settlements, had an IQR for 2010-2015 bounded by $Q_1 = 0.0083$ and $Q_3 = 0.0175$, while from 2015-2020, the IQR had $Q_1 = 0.005$ and $Q_3 = 0.0143$.

We used a t-test to compare the urban sprawl distributions of the three cities for the periods 2010-2015 and 2015-2020 to determine if there were significant differences between the two periods. Monterrey showed a remarkable change rate resulting in a t-value of 46.44 and p-value approaching zero. While in Mexico City, there is no significant change from one period to another, as shown in Fig. 5(b). In contrast, we found a substantial difference for Guadalajara with a t-value of 6.18 and p-value approaching zero for both periods. The results show that Monterrey and Guadalajara had significantly different urban sprawl distributions between 2010-2015 and 2015-2020.

V. DISCUSSION

Given the rapid growth of urban human settlements [3], and the pressing need to address climate change [8], extensive

(a) Mexico City (b) Guadalajara (c) Monterrey

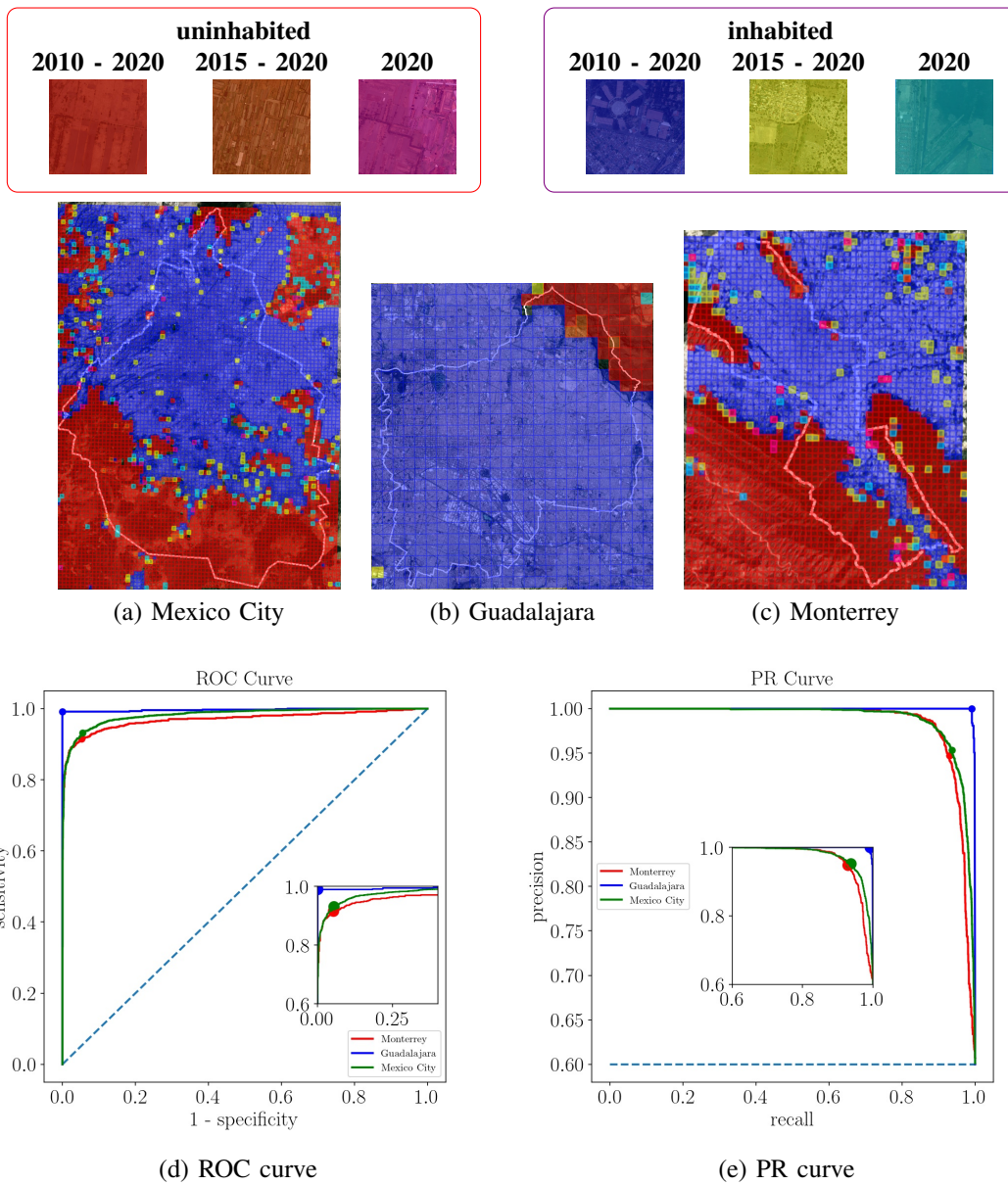(d) ROC curve

(e) PR curve

Fig. 4. Human Settlements Detection. To assess the detection performance, settlements were detected in the three largest cities (Mexico City (a), Guadalajara (b), and Monterrey (c)) using the EfficientNet-B7 classifier. The classifier was trained on the 2010 dataset and tested on information from images captured in years 2015 and 2020. The colored areas (red/blue) represent uninhabited/inhabited regions in 2010, 2015, and 2020. Yellow regions were uninhabited in 2010 but inhabited in 2015 and 2020, while orange regions were uninhabited in 2010 and 2015 but inhabited in 2020. Green areas were inhabited in 2010 but uninhabited in 2015 and 2020, and pink regions were inhabited in 2010 and 2015 but uninhabited in 2020. We evaluated the model's performance on unseen 2020 data from the three largest cities and plotted the performance curves (d) and (e).

research is being done to develop monitoring techniques to assess their impact. While some methods still rely on classical techniques, such as SVM and RF [13], [18], our approach follows the recent trend of employing deep learning techniques. However, in contrast to some recent studies [24], [25] that employ techniques with multiple remote sensing sources such as multispectral images, range, and radar, our approach uses a large dataset of multispectral images.

Our approach discriminates between inhabited and uninhabited classes without aiming to refine inferences into subclasses such as houses, roads, and buildings [16] and our architecture to semantically segment the area under analysis [23]

provided limited results. We rely on a publicly available, large dataset that systematically covers the entire country of Mexico, ensuring that data limitations do not constrain us as some researchers have been [20]. The related works, which employ other solution strategies, confirm that our results are outstanding within the state-of-the-art framework [13], [15], [16], [19], [20], [22].

To evaluate sprawl, most researchers follow a two-stage approach where the land usage problem is solved first, and then the sprawl is characterized. The human settlement detection stage is pursued using classical methods [37], [41], [43]–[46], [49], [50] (with techniques such as maximum likelihood,

(a) Settlements area

(b) Human settlements sprawl dynamics

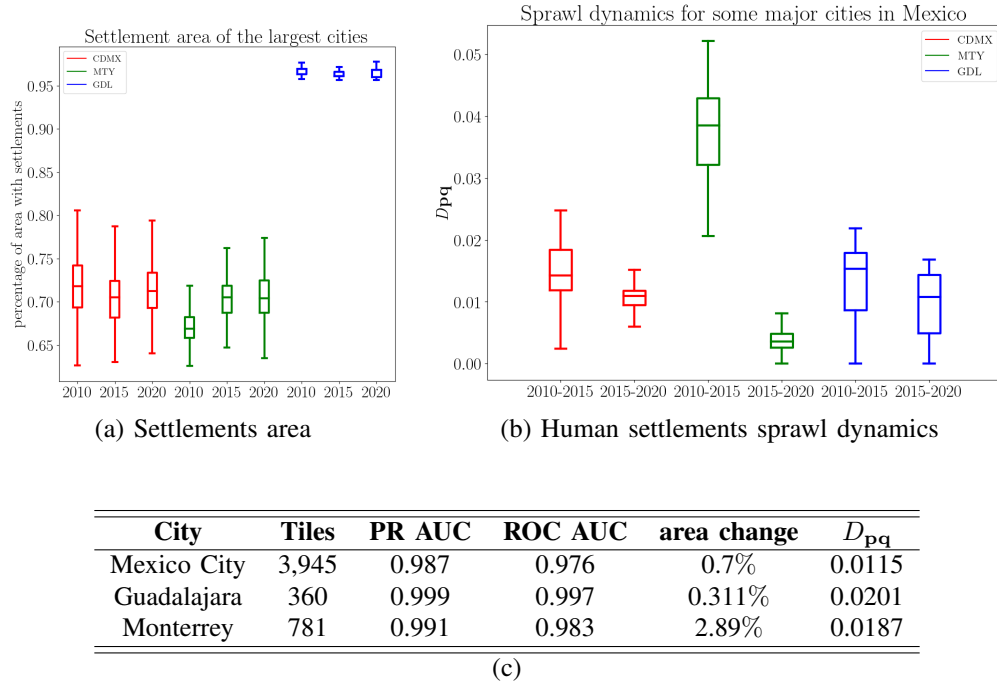| City	Tiles	PR AUC	ROC AUC	area change	$D_{\mathbf{pq}}$
Mexico City	3,945	0.987	0.976	0.7%	0.0115
Guadalajara	360	0.999	0.997	0.311%	0.0201
Monterrey	781	0.991	0.983	2.89%	0.0187

(c)

Fig. 5. Human settlement sprawl. Comparison of results from Guadalajara, Monterrey, and Mexico City. (a) Proportion of municipal territory occupied by human settlements. (b) Human settlement sprawl ($D_{\mathbf{pq}}$) for the top three Mexican cities. (c) Summary of results. Tiles correspond to 600×600 m^2 neighborhoods. Training is carried out on the 2010 dataset. The ROC AUC and PR AUC performance here is obtained with the 2020 dataset. Settlements sprawl is the difference from 2010 to 2020.

decision trees, SVM) or modern ones [16], [19]–[27], [34], [68], [69]. Regarding the human settlement problem, our research distances itself, offering a comprehensive evaluation of deep-learning-based methods and establishing a baseline for a dataset that is ample and covers most of Mexico. In the assessment of human settlements sprawl, researchers routinely employ the change of land use alone [37], [38], combined with other descriptors [39]–[42], or an ad hoc sprawl index, sometimes the result of multivariate analysis [35], [46], [47], [70]. These assessments describe aspects that may provide information about the features considered. Some researchers have used the Shannon entropy [49]–[51] as a principled, information-theory-based alternative. While appropriate to describe a snapshot, our approach fills the gap and enriches the characterization with an information-based description of the inter-period human settlement sprawl dynamics with the added value of being a metric.

In contrast to focused studies that examine niche aspects of human settlements like ancient sites [29], [30], rural areas [31], and informal settlements [32], [33], this paper presents a comprehensive analysis of human settlement dynamics across Mexico. By covering a country with diverse settlement scenarios, this study offers an inclusive and varied understanding of settlement patterns, developing strategies applicable to various countries with different geographical contexts.

This line of research will enable government planning agencies to make more frequent and detailed maps of urban sprawl with increased temporal resolution, bridging the gap between census years. These maps would allow more efficient monitoring of current sprawl trends, making predictive estimates about future urban development, and assessing the impact of past policy changes on sprawl dynamics, thereby aiding in sustainable urban planning and policy formulation.

VI. CONCLUSION

Given the rapid growth of human settlements, it is crucial to have effective mechanisms for monitoring and tracking their sprawl. This paper introduces a novel approach to assessing human settlement sprawl in Mexico using multispectral satellite image patches and Convolutional Neural Networks (CNN). By leveraging on an information-theory-based metrics, we achieve objectivity, standardization, and accuracy, which are characteristics that are further enhanced by the efficiency and predictive capabilities of deep learning. The performance attained in this task highlights the potential of this approach.

The study presents a methodology using remote sensing and deep learning to assess human settlements sprawl across Mexico. It utilizes and makes publicly available a large-scale dataset of labeled multispectral satellite images. The study's findings demonstrate the effectiveness of deep learning architectures, particularly EfficientNet. Future research will delve into ConvMixers as a promising, less complex alternative for detecting human settlements sprawl. This work sets a new standard in the use of deep learning and information-based metrics for large-scale human settlements analysis and may serve as a stepping stone for sustainable urban management.

Code and Data Availability Statement: Publicly available datasets were analyzed in this study. This data and code can be found at https://git.inegi.org. mx/laboratorio-de-ciencia-de-datos/human_settlements_ dynamics.

REFERENCES

[1] UN, "World Population Prospects 2019," United Nations, Tech. Rep., 2019, 10.1111/rssa.12104.

[2] V. Gaigbe-Togbe, L. Bassarsky, D. Gu, T. Spoorenberg, and L. Zeifman, "World Population Prospects 2022," United Nations, Tech. Rep., 2022.

[3] H. Ritchie and M. Roser, "Urbanization," *OWID*, 2018.

[4] Y. Bengio, Y. Lecun, and G. Hinton, "Deep learning for AI," *Commun. ACM*, vol. 64, no. 7, pp. 58–65, 2021, 10.1145/3448250.

[5] A. Sharif, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *CVPR*, 2014, pp. 806–813, 10.48550/arXiv.1403.6382.

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, 10.1038/nature14236.

[7] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean, "Carbon emissions and large neural network training," *arXiv:2104.10350*, 2021, 10.48550/arXiv.2104.10350.

[8] K. Anderson, B. Ryan, W. Sonntag, A. Kavvada, and L. Friedl, "Earth observation in service of the 2030 Agenda for Sustainable Development," *Geo-Spat. Inf. Sci.*, vol. 20, no. 2, pp. 77–96, 2017, 10.1080/10095020.2017.1333230.

[9] J. Reid, C. Zeng, and D. Wood, "Combining Social, Environmental and Design Models to Support the Sustainable Development Goals," in *IEEE Aerosp. Conf.* IEEE, 2019, pp. 1–13.

[10] E. C. Stokes and K. C. Seto, "Characterizing urban infrastructural transitions for the sustainable development goals using multi-temporal land, population, and nighttime light data," *Remote Sens. Environ.*, vol. 234, p. 111430, 2019.

[11] M. Prakash, S. Ramage, A. Kavvada, and S. Goodman, "Open earth observations for sustainable urban development," *Remote Sens.*, vol. 12, no. 10, p. 1646, 2020.

[12] A. L. Cowie, B. J. Orr, V. M. C. Sanchez, P. Chasek, N. D. Crossman, A. Erlewein, G. Louwagie, M. Maron, G. I. Metternicht, S. Minelli *et al.*, "Land in balance: The scientific conceptual framework for land degradation neutrality," *Environ. Sci. Policy*, vol. 79, pp. 25–35, 2018.

[13] M. Marconcini, A. Metz, S. Üreyen, D. Palacios, W. Hanke, F. Bachofer, J. Zeidler, T. Esch, N. Gorelick, and A. Kakarla, "Outlining where humans live, the World Settlement Footprint 2015," *Sci. Data*, vol. 7, no. 1, pp. 1–14, 2020, 10.6084/m9.figshare.12424970.

[14] J. Hoek and H. Friedrich, "Satellite-Based Human Settlement Datasets Inadequately Detect Refugee Settlements," *Remote Sens.*, vol. 13, no. 18, p. 3574, 2021, 10.3390/rs13183574.

[15] A. Rudiastuti, N. Farda, and D. Ramdani, "Mapping built-up land and settlements: a comparison of machine learning algorithms in Google Earth Engine," in *GSS*, vol. 12082. SPIE, 2021, pp. 42–52, 10.1117/12.2619493.

[16] C. Ayala, R. Sesma, C. Aranda, and M. Galar, "A Deep Learning Approach to Building Footprint and Road Detection in Satellite Imagery," *Remote Sens.*, vol. 13, no. 16, p. 3135, 2021, 10.3390/rs13163135.

[17] P. Gong, X. Li, and W. Zhang, "40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing," *Sci. Bull.*, vol. 64, no. 11, pp. 756–763, 2019, 10.1016/j.scib.2019.04.024.

[18] I. Tingzon, N. Dejito, R. Flores, R. Guzman, L. Carvajal, K. Erazo, I. Cala, J. Villaveces, D. Rubio, and R. Ghani, "Mapping New Informal Settlements using Machine Learning and Time Series Satellite Images: An Application in the Venezuelan Migration Crisis," in *ICAIG*. IEEE, 2020, pp. 198–203, 10.1109/AI4G50087.2020.9311041.

[19] C. Corbane, V. Syrris, F. Sabo, P. Politis, M. Melchiorri, M. Pesaresi, P. Soille, and T. Kemper, "Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery," *Neural. Comput. Appl.*, vol. 33, no. 12, pp. 6697–6720, 2021, 10.1007/s00521-020-05449-7.

[20] C. Qiu, M. Schmitt, C. Geiß, T. K. Chen, and X. Zhu, "A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully CNN," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 152–170, 2020, 10.1016/j.isprsjprs.2020.01.028.

[21] A. Rapuzzi, C. Nattero, R. Pelich, M. Chini, and P. Campanella, "CNN-Based Building Footprint Detection from Sentinel-1 SAR Imagery," in *IGRSS*. IEEE, 2020, pp. 1707–1710, 10.1109/IGARSS39084.2020.9323609.

[22] C. Qiu, M. Schmitt, H. Taubenböck, and X. Zhu, "Mapping Human Settlements with Multi-seasonal Sentinel-2 Imagery and Attention-based ResNeXt," in *JURSE*. IEEE, 2019, pp. 1–4, 10.1109/JURSE.2019.8809009.

[23] F. Wu, C. Wang, H. Zhang, J. Li, L. Li, W. Chen, and B. Zhang, "Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning," *Remote Sens. Environ.*, vol. 262, p. 112515, 2021, 10.1016/j.rse.2021.112515.

[24] G. Zitzlsberger, M. Podhorányi, V. Svatoň, M. Lazeckỳ, and J. Martinovič, "Neural Network-Based Urban Change Monitoring with Deep-Temporal Multispectral and SAR Remote Sensing Data," *Remote Sens.*, vol. 13, no. 15, p. 3000, 2021, 10.3390/rs13153000.

[25] S. Fibæk, C. Keßler, and J. Arsanjani, "A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning," *Int J Appl Earth Obs Geoinf*, vol. 105, p. 102628, 2021, 10.1016/j.jag.2021.102628.

[26] R. Ansari, R. Malhotra, and K. Buddhiraju, "Identifying Informal Settlements Using Contourlet Assisted Deep Learning," *Sensors*, vol. 20, no. 9, p. 2733, 2020, 10.3390/s20092733.

[27] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing," *Remote Sens.*, vol. 13, no. 15, p. 2965, 2021, 10.3390/rs13152965.

[28] R. Fan, R. Li, J. Li, W. Song, W. Han, J. Yan, and L. Wang, "Urban informal settlements classification via a transformer-based spatial-temporal fusion network using multimodal remote sensing and time-series human activity data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 111, p. 102831, 2022, 10.1016/j.jag.2022.102831.

[29] D. R. Osorio, S. Giraldo, E. Mazuera, A. Burbano, and E. Figueredo, "Beyond Visualization: Remote Sensing Applications in Prehispanic Settlements to Understand Ancient Anthropogenic Land Use and Occupation in the Sierra Nevada de Santa Marta, Colombia," *Latin American Antiquity*, pp. 1–21, 2023.

[30] N. Abate, D. Ronchi, V. Vitale, N. Masini, A. Angelini, F. Giuri, A. Minervino Amodio, A. M. Gennaro, and D. Ferdani, "Integrated Close Range Remote Sensing Techniques for Detecting, Documenting, and Interpreting Lost Medieval Settlements under Canopy: The Case of Altanum (RC, Italy)," *Land*, vol. 12, no. 2, p. 310, 2023.

[31] Z. Aamir, M. Seddouki, O. Himmy, M. Maanan, M. Tahiri, and H. Rhinane, "Rural Settlements Segmentation Based on Deep Learning U-Net Using Remote Sensing Images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 48, pp. 1–5, 2023.

[32] A. Crivellari, H. Wei, C. Wei, and Y. Shi, "Super-resolution GANs for upscaling unplanned urban settlements from remote sensing satellite imagery–the case of Chinese urban village detection," *International Journal of Digital Earth*, vol. 16, no. 1, pp. 2623–2643, 2023.

[33] K. G. Alrasheedi, A. Dewan, and A. El-Mowafy, "Using Local Knowledge and Remote Sensing in the Identification of Informal Settlements in Riyadh City, Saudi Arabia," *Remote Sensing*, vol. 15, no. 15, p. 3895, 2023.

[34] Z. Pan, J. Xu, Y. Guo, Y. Hu, and G. Wang, "Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net," *Remote Sens.*, vol. 12, no. 10, 2020, 10.3390/rs12101574.

[35] E. Gielen, G. Riutort, J. Miralles, and J. Palencia, "Cost assessment of urban sprawl on municipal services using hierarchical regression," *Environ. Plan. B: Urban Anal. City Sci.*, vol. 48, no. 2, pp. 280–297, 2021, 10.1177/2399808319869345.

[36] Y. Lin, T. Zhang, Q. Ye, J. Cai, C. Wu, A. Khirni Syed, and J. Li, "Long-term remote sensing monitoring on LUCC around Chaohu Lake with new information of algal bloom and flood submerging," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 102, p. 102413, 2021, 10.1016/j.jag.2021.102413.

[37] G. Hasnat, "Assessment of spatiotemporal distribution pattern of land surface temperature with incessant urban sprawl," *Environ. Chall.*, vol. 9, p. 100644, 2022, 10.1016/j.envc.2022.100644.

[38] M. Sridhar and R. Sathyanathan, "Assessment of Urban Expansion and Identification of Sprawl Through Delineation of Urban Core Boundary," *J. Landsc. Ecol.*, vol. 15, no. 3, pp. 102–120, 2022, 10.2478/jlecol-2022-0020.

[39] S. Lamichhane, "Assessment of urban sprawl and its impacts," Ph.D. dissertation, University of Salzburg, 2021.

[40] H. Ashraf, M. Mobeen, M. Miandad, M. Khan, G. Rahman, and S. Munawar, "Assessment of Urban Sprawl using Remotely Sense Data," *Ecol. Quest.*, vol. 33, no. 4, pp. 1–16, 2022, 10.12775/EQ.2022.030.

[41] B. Ashwathappa, M. Maddikeari, B. Das, R. Vishweshwaraiah, and R. Tangadagi, "Urban Sprawl Analysis and LULC change assessment in Bengaluru," *Res Sq*, 2022, 10.21203/rs.3.rs-1855333/v1.

[42] F. Cappelli, G. Guastella, and S. Pareglio, "Urban sprawl and air quality in european cities: empirical assessment," *Aestimum*, vol. 78, pp. 35–59, 2021, 10.2139/ssrn.3807084.

[43] S. Radhakrishnan and P. Geetha, "Urban Sprawl Assessment Using Remote Sensing and GIS Techniques," in *Intell. Sustainable Syst.* Springer, 2022, pp. 293–307, 10.21608/SJDFS.2022.269847.

[44] A. Ahmad, H. Gilani, S. A. Shirazi, H. R. Pourghasemi, and I. Shaukat, "Spatiotemporal urban sprawl and land resource assessment using Google Earth Engine platform in Lahore," *Comput. Earth Environ. Sci.*, pp. 137–150, 2022, 10.1016/B978-0-323-89861-4.00023-3.

[45] S. Nyongesa, M. Maghenda, and M. Siljander, "Assessment of urban sprawl, land use and land cover changes using remote sensing and landscape metrics," *J. Geogr. Environ. Earth Sci. Int.*, vol. 26, no. 4, pp. 50–61, 2022, 10.9734/jgeesi/2022/v26i430347.

[46] V. Chettry and M. Surawar, "Assessment of urban sprawl characteristics in Indian cities using remote sensing," *Environ. Dev. Sustain.*, vol. 23, no. 8, pp. 11 913–11 935, 2021, 10.1007/s10668-020-01149-3.

[47] J. Dey, S. Sakhre, R. Vijay, H. Bherwani, and R. Kumar, "Geospatial assessment of urban sprawl and landslide susceptibility," *Environ. Dev. Sustain.*, vol. 23, no. 3, pp. 3543–3561, 2021, 10.1007/s10668-020-00731-z.

[48] J. Bhattacharjee, S. Acharjee, and S. Mishra, "Identification of Urban Centre and Rural Growth Centres Around Guwahati and Its Surrounding Rural Region Using Hierarchical Settlements, Nested Hexagons, Remote Sensing and GIS," *Journal of Geomatics*, vol. 17, no. 2, pp. 198–210, 2023.

[49] S. Das and D. Angadi, "Assessment of urban sprawl using landscape metrics and Shannon's entropy model," *MESE*, vol. 7, pp. 1071–1095, 2021, 10.1007/s40808-020-00990-9.

[50] C. Stuht, "Las Vegas Metropolitan Area Urban Sprawl Assessment Using Shannon's Entropy," 2022, 10.1007/s40808-016-0209-4.

[51] N. Serdaroğlu Sağ, "Assessment of urban development pattern and urban sprawl using Shannon's entropy," 2021, 10.14687/jhs.v18i2.6158.

[52] R. Padmanaban, A. K. Bhowmik, P. Cabral, A. Zamyatin, O. Almegdadi, and S. Wang, "Modelling urban sprawl using remotely sensed data," *Entropy*, vol. 19, no. 4, p. 163, 2017, 10.3390/e19040163.

[53] D. Endres and J. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory.*, vol. 49, no. 7, pp. 1858–1860, 2003, 10.1109/TIT.2003.813506.

[54] D. Roberts, N. Mueller, and A. McIntyre, "High-dimensional pixel composites from earth observation time series," *IEEE Trans Geosci Remote Sens*, vol. 55, no. 11, pp. 6254–6264, 2017, 10.1109/TGRS.2017.2723896.

[55] INEGI, "Producción y publicación de la Geomediana Nacional a partir de las imágenes del Cubo de Datos Geoespaciales de México," INEGI, Tech. Rep., 2020.

[56] ——, "Censo de población y vivienda 2020: marco conceptual," INEGI, Tech. Rep., 2020.

[57] P. Merodio Gómez, O. J. Juarez Carrillo, M. Kuffer, D. R. Thomson, J. L. Olarte Quiroz, E. Villaseñor García, S. Vanhuysse, Á. Abascal, I. Oluoch, and M. Nagenborg, "Earth observations and statistics: Unlocking sociodemographic knowledge through the power of satellite images," *Sustainability*, vol. 13, no. 22, p. 12640, 2021, 10.3390/su132212640.

[58] A. Trockman and J. Zico, "Patches Are All You Need?" *arXiv*, p. 2201, 2022, 10.48550/arXiv.2201.09792.

[59] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *arXiv:2010.11929*, 2020.

[60] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, and J. Uszkoreit, "MLP-Mixer: An all-MLP Architecture for Vision," *NeurIPS*, vol. 34, 2021, 10.48550/arXiv.2105.01601.

[61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *CVPR*, 2016, pp. 770–778, 10.1109/CVPR.2016.90.

[62] I. Bello, W. Fedus, X. Du, E. Cubuk, A. Srinivas, T. Lin, J. Shlens, and B. Zoph, "Revisiting resnets: Improved training and scaling strategies," *NeurIPS*, vol. 34, pp. 22 614–22 627, 2021, 10.48550/arXiv.2103.07579.

[63] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for CNN," in *ICML*, 2019, pp. 6105–6114, 10.48550/arXiv.1905.11946.

[64] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241, 10.48550/arXiv.1505.04597.

[65] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Eff-UNet: Novel Architecture for Semantic Segmentation in Unstructured Environment," in *CVPRW*, 2020, pp. 1473–1481, 10.1109/CVPRW50498.2020.00187.

[66] C. Manning and H. Schutze, *Foundations of statistical natural language processing*. MIT press, 1999, 10.1145/601858.601867.

[67] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *CVPR*. IEEE, 2009, pp. 248–255, 10.1109/CVPR.2009.5206848.

[68] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *CVPR*, 2017, pp. 1492–1500, 10.1109/CVPR.2017.634.

[69] M. R. Ibrahim, H. Titheridge, T. Cheng, and J. Haworth, "predict-SLUMS: A new model for identifying and predicting informal settlements and slums in cities from street intersections using machine learning," *Comput. Environ. Urban Syst.*, vol. 76, pp. 31–56, 2019, 10.1016/j.compenvurbsys.2019.03.005.

[70] S. Al Saleh, R. Abu Samra, T. Hegazy, M. Abd, and M. Mohamed, "Urban sprawl assessment and modelling of Shahat using space data," *SJDFS*, vol. 12, no. 1, pp. 50–55, 2022, 10.21608/SJDFS.2022.269847.

Antonio Briseño is a software engineer pursuing a master's degree in computer vision at Instituto Politécnico Nacional. He has experience developing and leading various web and mobile applications and ETL systems in logistics, retail stores, and education. Also, He has expertise in scaling and portability using cloud services software. He has participated in various hackathons in the IT industry. Currently, his work focuses on machine learning algorithms, image analysis, and Geographical Information Systems applications.

Joaquin Salas is a professor in the field of Computer Vision at Instituto Politécnico Nacional. Member of the Mexican National Research System, his research interests include monitoring natural systems using visual perception and aerial platforms. Salas received a Ph.D. in computer science from ITESM, México. He has been a visiting scholar at MIT, Stanford, Duke, Oregon State, Xerox PARC, the Computer Vision Center, and the ÉNST de Bretagne. He has served as co-chairperson of the Mexican Conference for Pattern Recognition three times. Salas was a Fulbright scholar for the US State Department. He has been invited as editor for Elsevier Pattern Recognition and Pattern Recognition Letters. For his services at the Instituto Politécnico Nacional, he received the *Lázaro Cárdenas* medal from the President of Mexico.

Elio Villaseñor is a data scientist holding a Master's degree in Mathematics and a Ph.D. in Computer Science awarded by the National Autonomous University of Mexico. With extensive experience in education, he has taught mathematics from elementary school to postgraduate programs. As a dedicated researcher, he primarily focuses on machine learning, neural networks, and data science applications with significant societal implications. Presently, Elio serves as the lead scientist at the Data Science Lab of the National Institute of Statistics and Geography in Mexico. Additionally, he is a member of the United Nations' Expert Committee on Big Data for Official Statistics.

Ranyart R. Suarez is a data scientist with a Ph.D. in Computer Science from the Universidad Michoacana de San Nicolás de Hidalgo. He currently works at the Data Science Laboratory of the National Institute of Statistics and Geography (INEGI). His areas of study include machine learning, deep learning, and remote sensing. Alternatively, he also has experience in classrooms teaching graduate-level courses focused on the machine learning field. Currently, his research activities are focused on using machine learning techniques and geospatial information to develop alternative methodologies that approximate the calculation of the Sustainable Development Goals (SDGs) indicators proposed by the United Nations.

Danielle Wood is an Assistant Professor in the Media Arts & Sciences Program and holds a joint appointment in the Department of Aeronautics & Astronautics at the Massachusetts Institute of Technology. Within the Media Lab, Prof. Wood leads the Space Enabled Research Group, which seeks to advance justice in Earth's complex systems using space-enabled designs. Prof. Wood is a scholar of societal development with a background in satellite design, earth science applications, systems engineering, and technology policy. In her research, Prof. Wood applies these skills to design innovative systems that harness space technology to address development challenges worldwide. Before serving as MIT faculty, Professor Wood held positions at NASA Headquarters, NASA Goddard Space Flight Center, Aerospace Corporation, Johns Hopkins University, and the United Nations Office of Outer Space Affairs. Prof. Wood studied at the Massachusetts Institute of Technology, earning a PhD in engineering systems, SM in aeronautics and astronautics, SM in technology policy, and SB in aerospace engineering.