Future Urban Development

Leveraging AI for Sustainable Decisions

RISE Report 2025:99

2025-10-31

Parametric

RISE

Abstract

The project explored how artificial intelligence (AI), combined with synthetic datasets and rule-based models, can support decision-making in early-stage urban development. Using the generative design platform Hektar as a test bed, the team developed, implemented, and evaluated two complementary computational approaches: a deterministic, explainable algorithm and machine-learning models trained on large-scale synthetic data representing over two million urban plot configurations.

The deterministic model demonstrated high precision, achieving less than ±5 % deviation between user-defined targets (FAR, SCR) and resulting outputs for 95 % of test plots. The machine-learning work progressed in two stages. In the first stage, a numerical Multilayer Perceptron (MLP) outperformed a convolutional neural network (CNN) in predicting Site Coverage Ratio (SCR) from compact geometric descriptors. In the second stage, a refined model predicted probability distributions of SCR outcomes, reflecting the stochastic generation of building configurations in the updated Hektar system. Together, these methods established a reproducible workflow that translates user goals into valid spatial outcomes while introducing probabilistic reasoning to early-stage planning.

The project demonstrates that small, task-specific AI models can be computationally efficient while delivering substantial benefits. By improving the precision of density and form assessments, such models can contribute to reduced material use, more efficient land allocation, and lower climate impact in the built environment.

All predictive models, datasets, and documentation are published openly on GitHub to support further research and industry adoption. By shifting from form generates data to data generates form, the project outlines a scalable pathway toward prescriptive, data-driven urban planning tools capable of supporting more sustainable, evidence-based decisions.

Project partners

RISE Research Institutes of Sweden AB

- Alexander Gösta
- Leon Sütfeld

Parametric Solutions AB

- Sofia Malmsten
- Erik Forsberg
- Martin Skarby
- Alexander Radne
- Simon Brobäck

Reference group:

- EttElva arkitekter
- AFRY
- Elisabeth Flakierska, senior adviser
- Camilla Berggren-Tarrodi, RISE

Contact

Alexander Gösta, alexander.gosta@ri.se

Funding programme

Vinnova – Utlysning Avancerad och innovativ digitalisering 2024 – ettåriga projekt

Publication

2025-10-31

RISE Rapport 2025:99

ISBN 978-91-90036-88-4

Contents

Introduction

Background

Cities face increasing urbanisation, which places significant pressure on land use, infrastructure, and housing supply. Urban development must balance economic, social, and environmental factors to meet the need for sustainable and attractive urban environments, with challenges such as noise, daylight conditions, access to services, transportation, and climate impact.

Early-stage property development is often time-consuming and has access only to fragmented data, which makes decision-making processes complex and high-risk. Artificial intelligence (AI) shows great potential to accelerate and support decisions in the field.

Purpose

The project aimed to explore and develop AI-based methods to support urban development and property planning decision-making, focusing on sustainability and feasibility.

Objectives

The project formulated the following objectives:

- Develop an operational system prototype at TRL 7 (Technology readiness level) for exploring AI-based methods for property development.
- Produce conference contributions relevant to the industry.
- Test and validate synthetic data as training material for AI models.

Methodology

The team used the existing Hektar platform developed by Parametric Solutions as the foundation for the work.

The first step was identifying measurable parameters that could translate complex urban design questions into quantitative metrics.

Hektar was used as an engine for generating synthetic data and as a test environment for integrating and validating the developed methods. These datasets were enriched with metadata and statistical descriptors, enabling training and testing of different models.

The model development process followed two parallel tracks:

- A deterministic rule-based algorithm that generated and ranked design solutions based on a typology library.
- Machine learning models are trained to rapidly predict likely outcomes based on the plot's shape and basic attributes.

The workflow was iterative: the models were trained and refined through cycles of computation and simulation. Repeated testing and verification in workshops involving the project team and reference group stakeholders produced qualitative feedback.

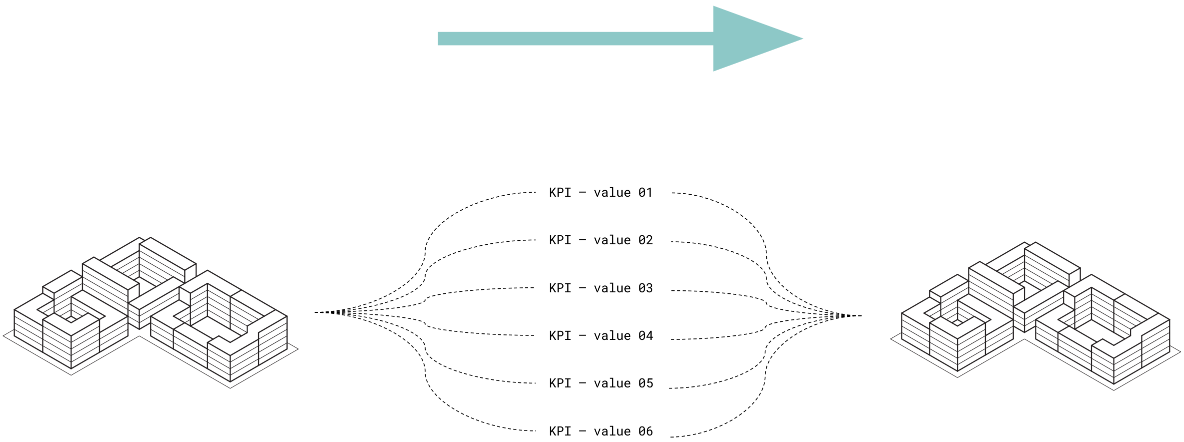

Image 1. *The project aims to develop a digital workflow that moves from descriptive to prescriptive design. Users can define goals and priorities, generate and evaluate multiple design alternatives, and use data-driven feedback to inform decisions at early planning stages.*

Implementation

Identifying measurable parameters

The first step in developing the computational methods was to determine which aspects of urban form we could translate to measurable, comparable, and optimisable parameters. Sustainable urban development depends on a complex interaction of social, ecological, and economic factors, many of them qualitative and context-specific. To build a clear and testable foundation for modelling, the team focused on urban density as the primary variable for exploration and optimisation.

Density provides a measurable bridge between physical form and broader sustainability outcomes. Density is not inherently positive or negative considering sustainability goals, but influences multiple dimensions of sustainability. Higher density can support efficient land use, public transport viability, and urban vitality, while lower density may improve daylight access, thermal comfort, and biodiversity. Because of this duality, density is a good metric for examining the spatial and environmental trade-offs embedded in planning decisions.

Within this broader framework, the project selected the Site Coverage Ratio (SCR) as the controllable metric for optimisation. Whereas the Floor Area Ratio (FAR) describes overall built volume, it combines SCR and building height that are partly predetermined by contextual and regulatory constraints. On the other hand, SCR directly expresses how much of a site's surface is occupied by buildings and can therefore be influenced in the early design phase. It provides a practical way to test and compare feasible development intensities while maintaining height, typology, and programme distribution flexibility.

We selected the Site Coverage Ratio (SCR) as the controllable metric for optimisation. Whereas the Floor Area Ratio (FAR) describes overall built volume, it combines SCR and building height.

Desired to maximise while maintaining other values

given by context and preference

$$FAR = \frac{\overbrace{SCR \times avg.floors}^{GFA}}{Site\ Area}$$

given by context and preference

Fixed for each site

FAR = Floor area ratio
GFA = Gross floor area
SCR = Site coverage ratio

Image 2. *Controllable metric in the design optimisation process.*

Contextual and regulatory constraints partly predetermine height. SCR directly expresses how much of a site's surface is occupied by buildings and can therefore be influenced in the early design phase. It provides a practical way to test and compare feasible development intensities while maintaining height, typology, and programme distribution flexibility.

The approach allowed for systematic experimentation with density as a measurable concept that connects the algorithmic optimisation process with the broader goals of sustainable and responsive urban design.

Since FAR is calculated as SCR multiplied by the average number of floors, various combinations of footprint and height can yield the same FAR. A site with FAR 2.0 could be realised as SCR 0.25 with eight storeys (slender slabs or towers), SCR 0.50 with four storeys (perimeter blocks), or SCR 0.67 with three storeys (deep, low buildings). These variants differ morphologically. They vary in frontage and street enclosure, courtyard size and permeability, daylight access, microclimate, and the balance between public and private open space. While FAR alone cannot reveal how a city will look or feel, SCR directly governs ground occupation and thus defines urban form at eye level: block depth and spacing, continuity of street edges, and the size and configuration of open areas.

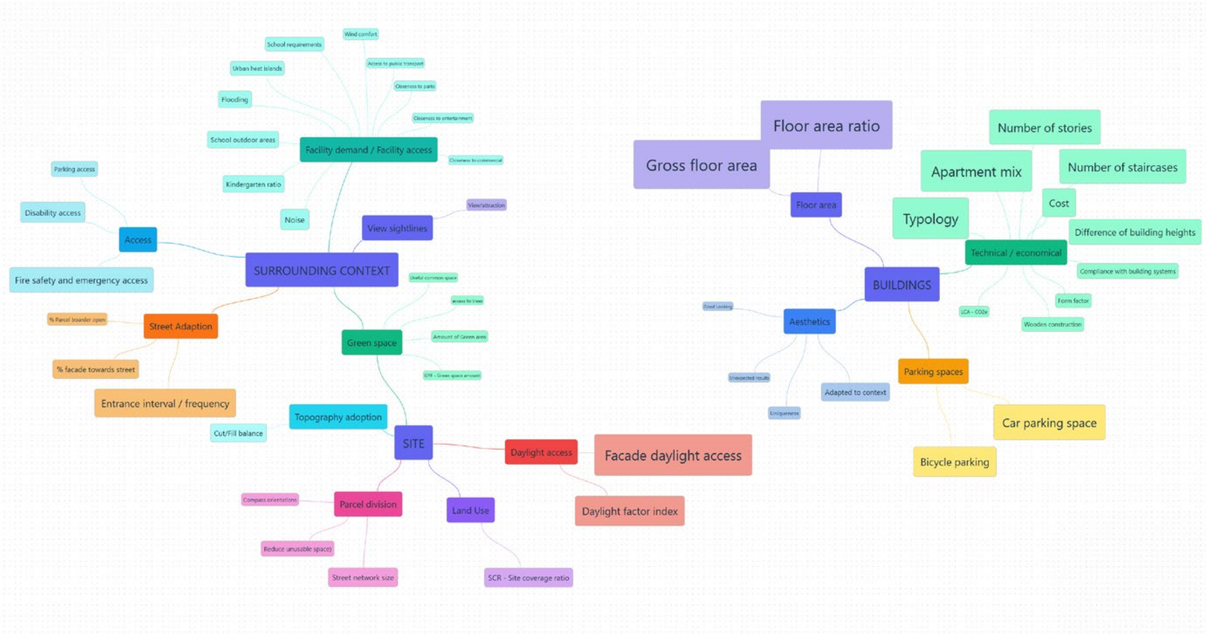

Image 3. *Overview of measurable metrics in urban planning and real estate development.*

The diagram shows how quantitative indicators describe both physical and contextual aspects of urban form, grouped into three domains: Buildings, Site, and Surrounding context. Building metrics include FAR, floor area, typology, height, and material or performance qualities. Site factors cover parcel division, street network, land use, SCR, daylight, and topography. Contextual parameters describe accessibility and comfort, such as proximity to transport, green space, parking, fire access, and views.

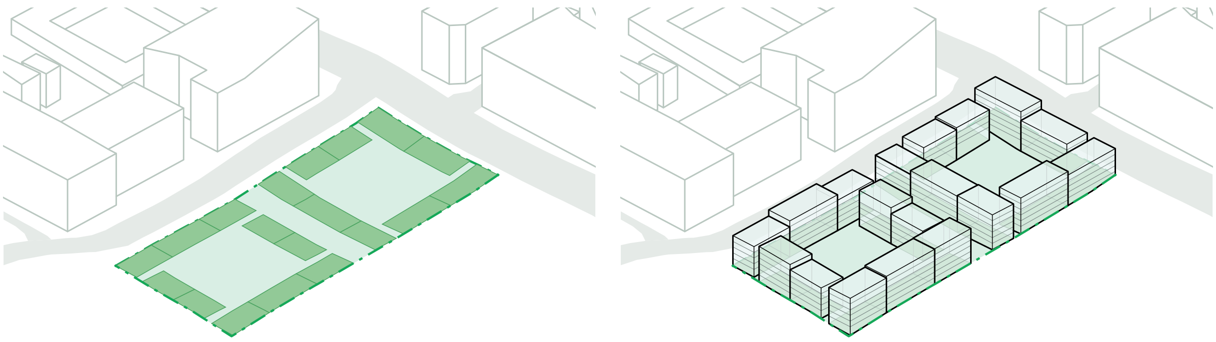

Image 4. *Illustration of spatial density metrics*

The Site Coverage Ratio (SCR) expresses the share of the site covered by buildings, the Gross Floor Area (GFA) measures total built floor space, and the Floor Area Ratio (FAR) relates total floor area to site area, together describing different dimensions of urban density.

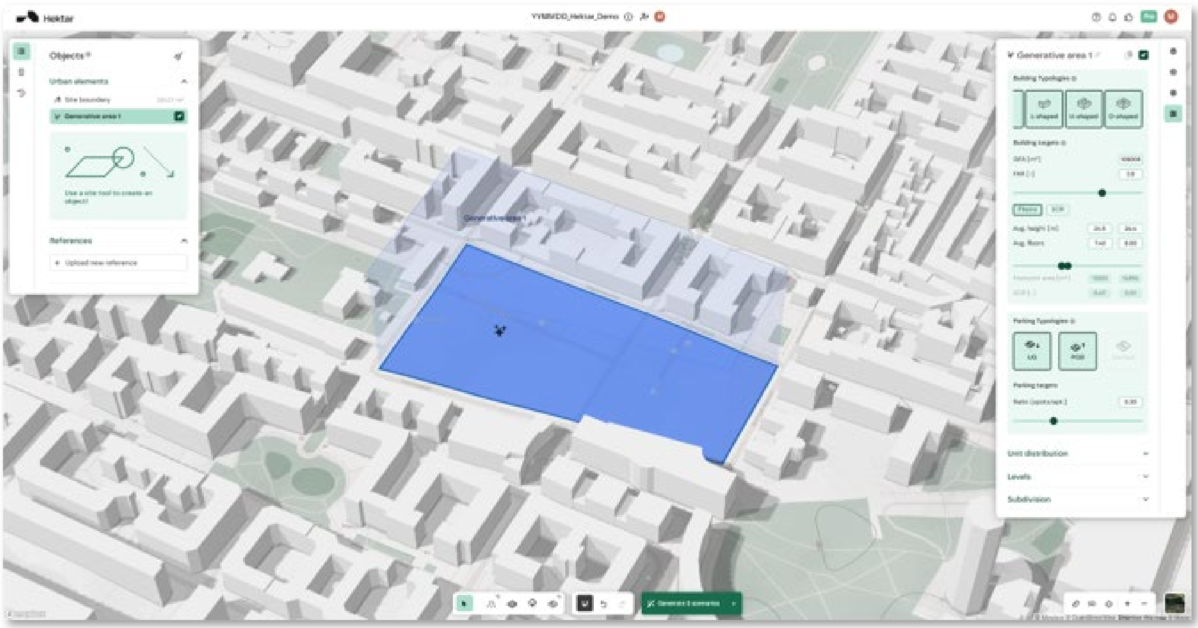

Image 5. *A generative area in Hektar, a plot where geometry is created by the software.*

How Hektar works

Hektar is a web-based, augmented CAD environment developed by Parametric Solutions.

It is designed for large-scale urban form simulation and analysis, providing manual and generative modes for creating and evaluating site configurations. The tool combines geometric editing, procedural generation, and real-time calculation of urban density metrics.

The platform supports conventional drawing augmented by continuously providing analytical feedback. Geometry can be adjusted incrementally, with real-time updates of key metrics such as Floor Area Ratio (FAR), Site Coverage Ratio (SCR), and height. The system also incorporates contextual logic to auto-complete shapes or extend edges while giving the designer complete control over geometry and parameters. This direct link between traditional design interaction and computational analysis makes testing and refining design hypotheses step by step possible.

Hektar provides a feature for procedural generation of site layouts. Users can set specific input parameters, such as target density, building typology, or the number of floors. The system then assembles various configurations using a predefined library of building types and shapes. This allows for rapid scenario testing, enabling the generation and comparison of multiple layouts based on consistent metrics. Importantly, the workflow is reversible; geometry and metrics can influence each other depending on the design stage.

Each generated element site, parcel, or building stores a consistent set of geometric and contextual attributes and computed indicators such as FAR and SCR. Hektar provides a design environment that links form and performance through integrating data overlays and generative logic.

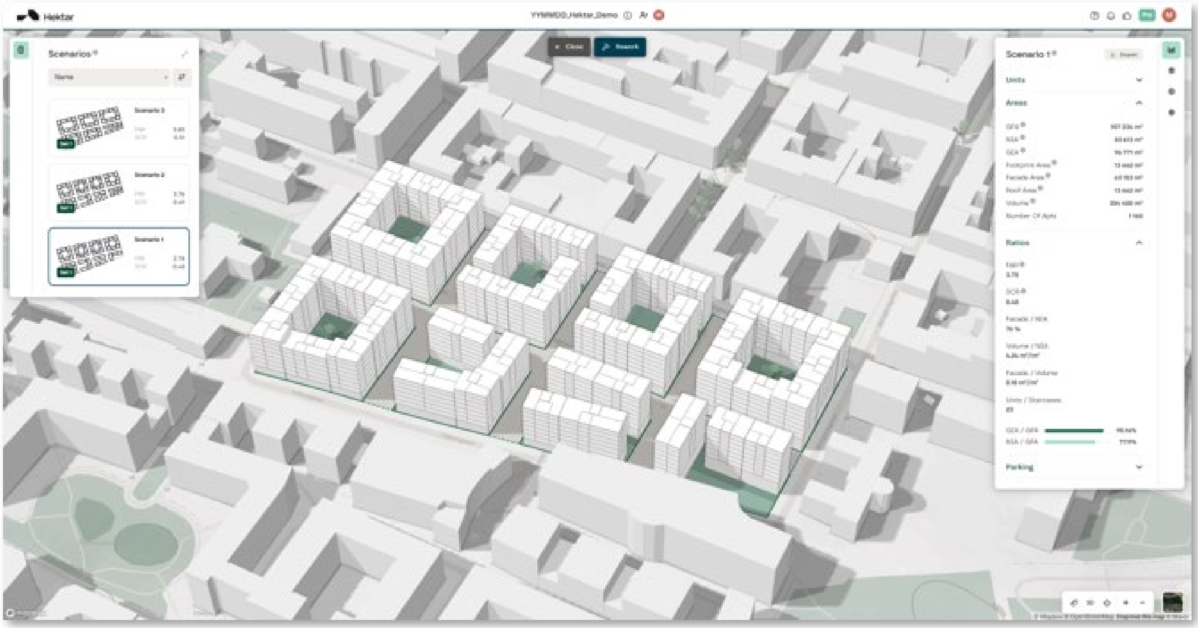

Image 6. *A finished proposal from Hektar. In this case, three different configurations are created.*

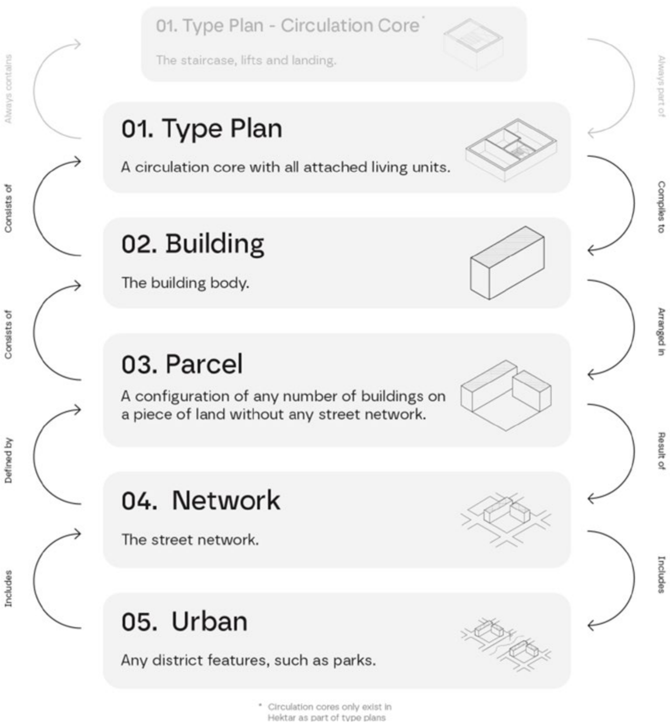

Image 7. *Hektar organizes spatial data through a nested hierarchy (1) Type Plan – circulation core with attached units (2) Building – assembled from one or more type plans (3) Parcel – one or several buildings within a site boundary (4) Network – the connecting street system (5) Urban – aggregated areas such as districts or parks. Each level is parametrically linked to the next, allowing modifications at one scale to propagate through the system.*

Generating synthetic training data

The training data used in this project was generated through the Hektar platform. Within Hektar, the team created synthetic site models combining algorithmically generated and real-world geometries. Each site was subdivided into parcels, each defined by its polygonal boundary and associated metadata such as perimeter, area, aspect ratio, and side count. The JSON structure of a typical site file contains a SiteBoundary and a list of Parcels, each described by a sequence of X–Y coordinates.

To train and test the prediction models, tens of thousands of such parcels were generated, covering a wide range of plan shapes and side counts from simple four-sided lots to irregular polygons. For each parcel, multiple building typologies were instantiated, systematically varying parameters such as building footprint type (L, U, slab, perimeter block), number of volumes, and number of floors. The typologies were combined in different building configurations (e.g., single block, double slab, perimeter block), resulting in diverse physical outcomes even for the same parcel geometry.

Each generated case was annotated with quantitative indicators describing density and form such as FAR (Floor Area Ratio), SCR (Site Coverage Ratio), building height, envelope area, and volumetric ratios alongside contextual descriptors like orientation, adjacency, and access edges. This multi-layered data structure created a comprehensive synthetic dataset suitable for supervised learning, where input features (plot geometry, typology, and context) could be directly mapped to output targets (e.g., predicted FAR and SCR).

These datasets established a controlled but highly diverse training environment, enabling the model to generalise across a wide range of urban morphologies while remaining interpretable and reproducible.

Property	Value
Created	2025-05-19
Creation time	6 h 18 m
Rows	2 290 945
Sites (s)	886
Road types / widths (t)	5
Divisions (d)	1
Parcels (p)	Varying (see .csv)
Typologies (t)	6
Footprints (f)	16

Table 1. *Dataset 2 - synthetic data*

Purpose: *Large-scale simulation of urban form for FAR / SCR prediction*

Composition: *Each site subdivided into parcels with polygonal boundaries and metadata (area, perimeter, aspect ratio, side count, etc.)*

Variations: *Tens of thousands of combinations of parcel shapes and building typologies (L-, U-, O-, corner-, slab-, and point-house configurations)*

Outputs: *Each case annotated with quantitative indicators FAR, SCR, height, envelope area, volumetric ratios and contextual descriptors such as orientation, adjacency, and access edges.*

Columns included	Description
SiteID, ParcelID	*Unique identifiers for each site / parcel*
ParcelArea, Perimeter, AspectRatio, SideCount	*Geometric properties of each parcel*
BuildingTypeGroup	*Selected building type: PointHouse, SlabBuilding, CornerBuilding, LShapedBuilding, UShapedBuilding, OShapedBuilding*
ParcelTypology	*ID of typology assigned to parcel (may be None)*
ParcelTypologyHasPointHouse, HasSlabBuilding, HasLShapedBuilding, HasUShapedBuilding, HasOShapedBuilding, HasCornerBuilding	*Boolean flags for each building subtype present*
NumFloors, VolumeCount, FootprintIndex	*Parameters defining the instantiated building*
FAR, SCR, BuildingHeight, EnvelopeArea, VolumeToAreaRatio	*Calculated performance metrics*
Orientation, AdjacencyCount, AccessEdgeCount	*Contextual indicators of site configuration*

Table 2. *Dataset 2 - columns included*

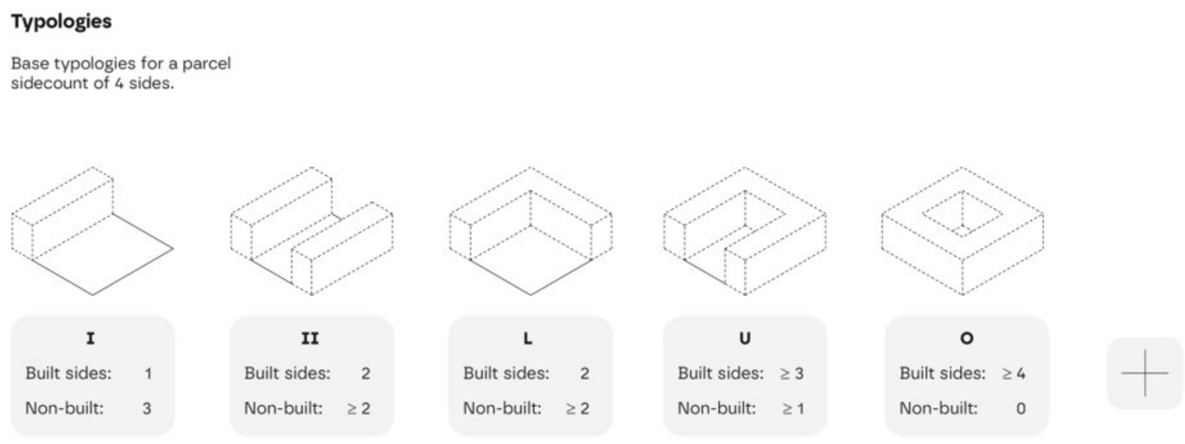

Image 8. *Building typologies in Hektar*

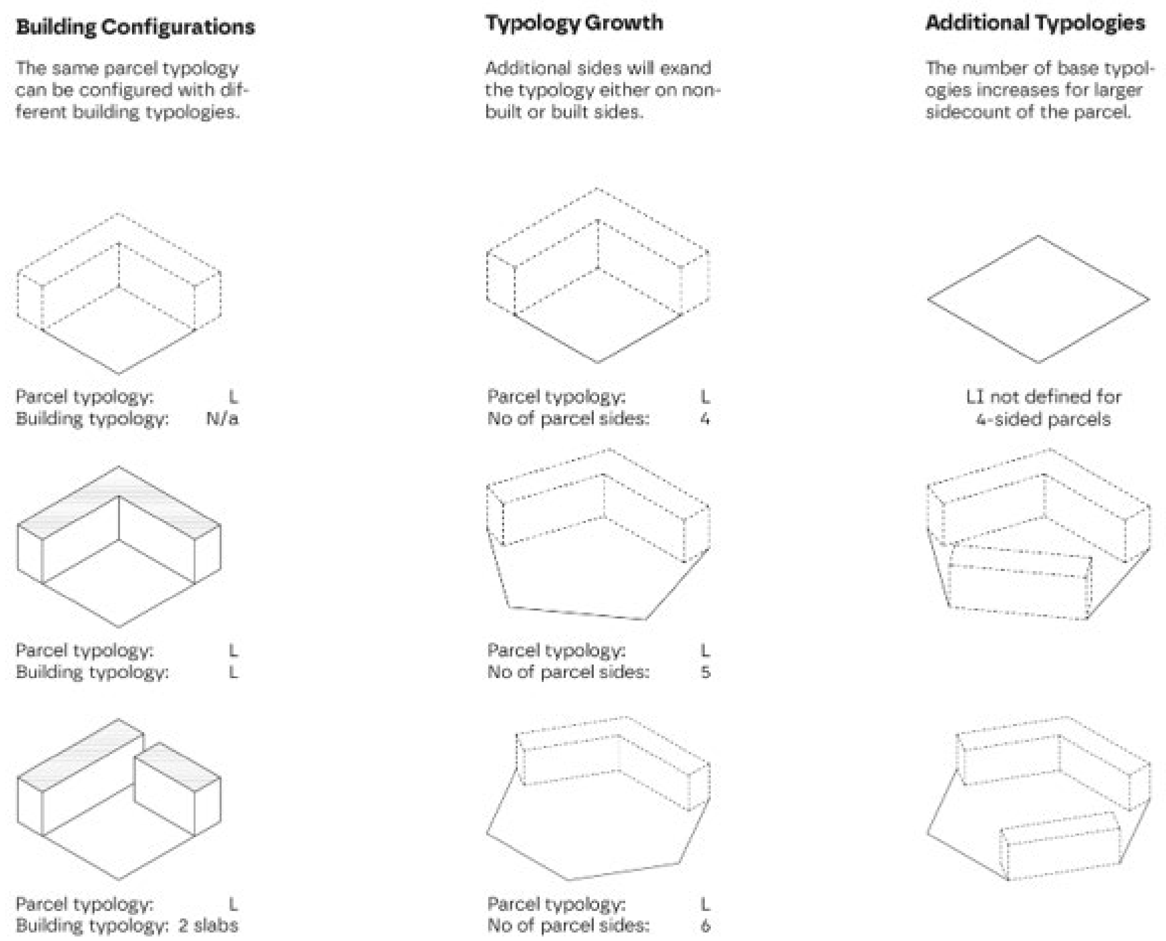

Image 9. *Building Configurations: The same parcel typology can host multiple building configurations, such as single or double slabs.*
Typology Growth: The number of parcel sides allows expansion of base typologies e.g., a four-sided parcel supports fewer variants than a six-sided one.
Additional Typologies: As side count increases, the possible base typologies grow combinatorially, forming a richer dataset of geometric conditions.

Deterministic algorithm

Producing the training data for the AI models led to new insights on improving calculation time and the quality of the existing rule-based sequence in Hektar. Producing the numerical measurements to categorise plots also led to possible improvements on the regular rule-based system. These improvements were made and implemented in Hektar's live server.

The user can specify desired metrics in the new system, such as Floor Area Ratio (FAR), Site Coverage Ratio (SCR), or height limits. These targets define the boundaries within which the algorithm operates. A tagged library of building typologies is loaded, each represented by a structured data record rather than explicit geometry, similar to the training data for the AI. Every typology contains precomputed attributes describing its geometry, performance, and contextual compatibility such as footprint type, number of floors, yield, coverage, daylight proxy, and access edges.

Before generation, the model filters the library to retain only typologies that can physically fit within the selected parcel. This step checks geometric feasibility against boundary shape, frontage, and setback parameters.

Feasible typologies are then systematically combined into candidate configurations. The algorithm calculates each combination's resulting FAR, SCR, and related indicators. Options are compared against the target values and ranked according to how closely they match the input criteria. The model is fully deterministic: identical inputs always yield the same ranked outcome, ensuring reproducibility and interpretability.

Once the algorithm identifies the top-ranked configuration, geometry is generated only for the top-performing alternatives. The results are displayed together with key performance indicators, enabling direct comparison between target and achieved values.

The implementation reached full operational functionality within the Hektar platform. The deterministic model now enables real-time prediction of density outcomes and validated geometry generation under defined constraints. It also provides a reliable reference for evaluating machine-learning approaches developed in parallel, establishing a benchmark for explainable and reproducible prediction of urban form.

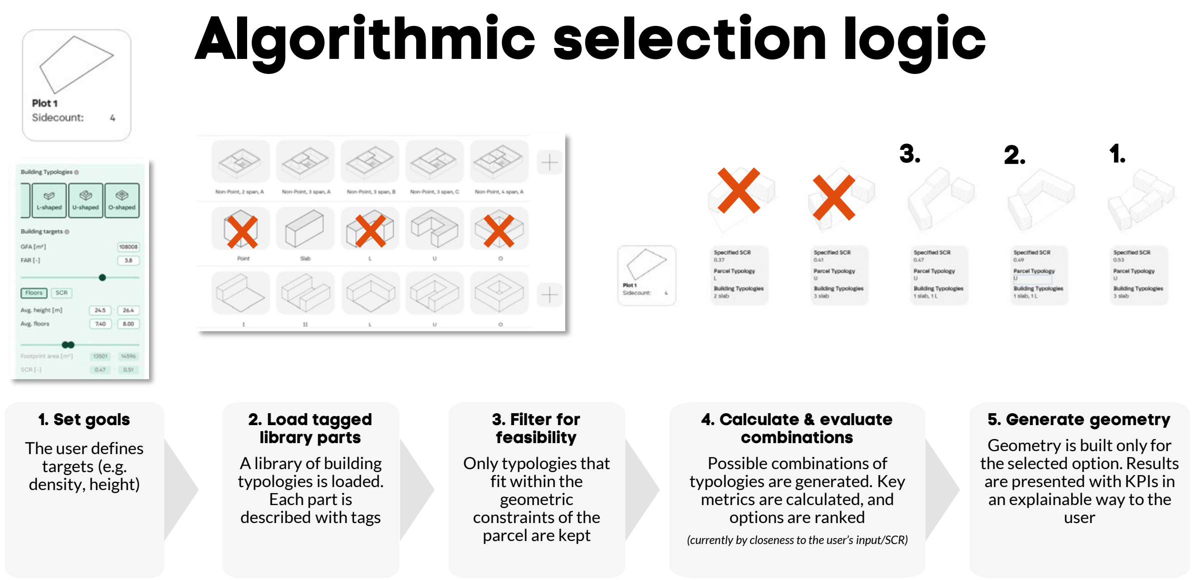

Image 10. *Algorithmic selection logic in the rule-based model.*

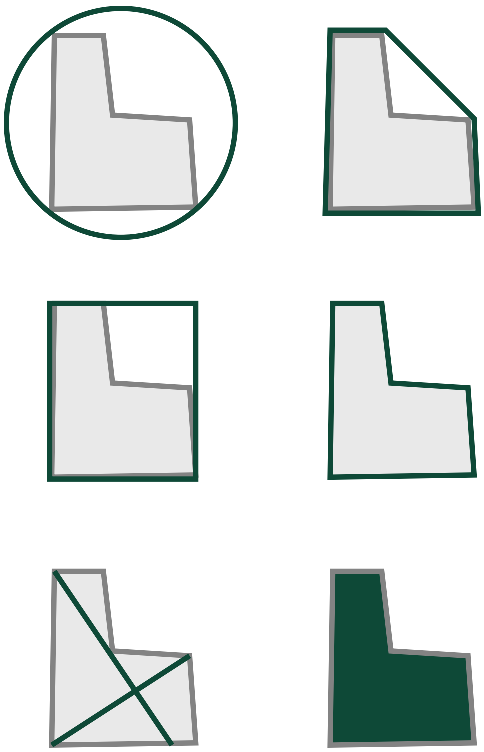

Image 11. *Illustration of the Numerical parameters for categorising plots, explainer in the table to the right.*

Parameter	Description
Area (normalized)	*total surface area relative to the dataset.*
Circumference (normalized)	*total boundary length relative to area.*
Ratio of axes	*relation between the site's major and minor axis lengths.*
Absolute length of axes (normalized)	*overall size and elongation of the site.*
Solidity	*ratio of site area to convex hull area, describing compactness.*
Rectangularity	*ratio between site area and its smallest bounding rectangle, indicating how rectangular or irregular the shape is.*
Circularity	*ratio between site area and its smallest enclosing circle, measuring how close the shape is to circular.*
Moment of inertia	*distribution of area relative to the centroid, capturing how mass or form is spread around the center.*

Table 3. *Parameters for the numerical approach.*

Testing machine learning methods

The compute requirements for Hektar usually result in wait times of around 20 seconds for the user. While this may be acceptable in many cases, previewing key metrics of the final construction plans early would be better, as it would enable the user to quickly test different parameters and see their impact on important metrics. For example, the selected building type (quarters, point houses, L-shaped buildings, etc.) significantly affects the Site Coverage Ratio (SCR), which shows how much of the site is covered by buildings and how many square metres of usable floor space will be generated per floor. Machine learning (ML) can help provide fast estimates of key metrics such as SCR. With this project, we aimed to create a proof of concept and explore different ways of representing sites for ML models, as well as options for the models themselves.

The main goals for the ML part of the project were:

- Train an AI model to quickly predict SCR statistics for a site based on its shape and typology
- Support early design decisions by showing likely development outcomes without needing full geometry generation

Address research questions:

- How to best represent site parameters such as shape and size within an ML model?
- Which ML models are most suitable for this task?

Although the overall objective – predicting SCR from site parameters – remained consistent during the project, the ongoing development of Hektar meant that the specific composition of datasets and thus the task for the ML models shifted over time, resulting in a two-stage process. These stages will be described separately in the following text.

Site Representations

In Hektar, each site is defined as a 2D shape with arbitrarily many vertices (corners) connected with straight lines. This information can be represented to neural networks and other ML methods in multiple ways. Graph neural networks (GNNs) and transformer architectures, for instance, should generally be suited to handle variable-length lists of vertex coordinates. Architectures for image processing, such as convolutional neural networks (CNNs) and vision transformers could work with pixel-based renders of the sites' top-down shapes. And finally, Multilayer-Perceptrons (MLPs), as well as other classical ML methods (gradient boosting methods, random decision forests) typically work best with fixed-length numerical representations.

Which of these work best for any given task cannot be determined a priori, but needs to be tested empirically. Given the proof-of-concept scope of the project and the often large amounts of data required to train GNNs and transformer architectures, we decided to focus on image-based and fixed-length numerical representations, and thus discarded variable-length vertex coordinates lists.

Image-based representation

Top-down renders of the site's shape keep all information about a site's shape alive, albeit at a somewhat coarse resolution due to fixed dimensions of the input images (e.g., 512x512). One challenge with this representation style for the task at hand is the highly variable overall size of a site; Hektar allows for the planning of very small patches of land, e.g., in dense inner-city areas, as well as for multiple hectare spanning developments in the countryside. We therefore took the decision to scale each site to get close to the outside boundaries of the fixed-sized canvas (512x512) and include the information about the scaling via a second channel. To keep all inputs in the image domain, we opted to represent a site's scaling via a checkerboard in a separate color channel; the larger the site the smaller the checkerboard's squares and vice versa. An alternative approach would be to include information about scaling and potentially additional factors as numerical inputs at a later layer of the network. However, given the scope of the project we decided to prioritize other aspects in the project.

Numerical Representations

To create fixed-length numerical representations of a site's shape, we computed several descriptive metrics per site (see image 11 and table 3).

While the number of metrics that can describe a site is vast, this selection should describe the site shapes and their variability within the context of urban planning. However, follow-up work should further investigate the usefulness of additional numerical shape descriptors.

Parcels and Streets

Large sites without any subdivision would result in difficulties reaching the innermost buildings and places from the nearest street. Hence, Hektar automatically subdivides sites into individual parcels for large sites. A parcel is thus a continuous area on which buildings can be constructed and may be separated from other parcels on the same site by streets.

The width of the streets that subdivide a site into individual parcels is a user-determined parameter and an important variable to consider in predicting, e.g. a site's resulting SCR. Wider streets take up more of the site's overall area and thus tend to lower resulting SCR figures.

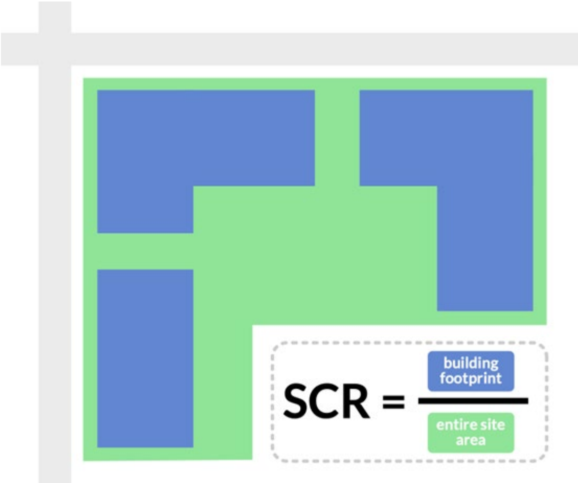

Image 12. *Visual explanation of the Site Coverage Ratio (SCR).*

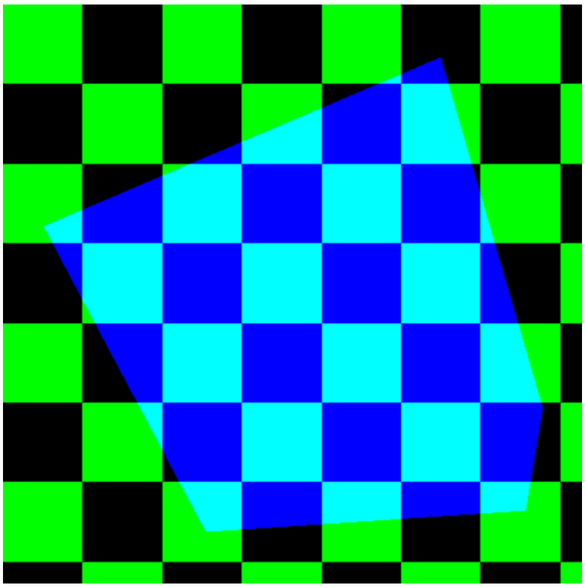

Image 13. *Image-based representation of a site's shape (blue channel) and scale (green channel).*

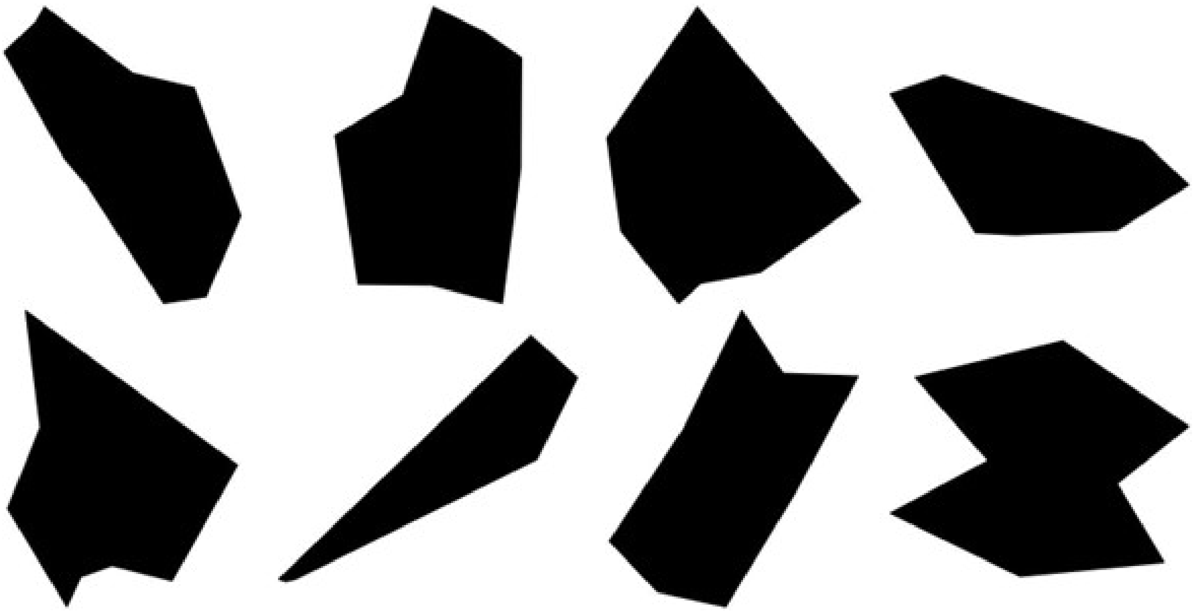

Image 14. *Examples of synthetically generated irregular site shapes (not representative of the distribution of site shapes used in the study).*

Site Shapes

We opted not to use real-world sites to select sites to train the algorithm. This decision was based on the observation that the distribution of shapes in the real world is skewed towards regular and rectangular shapes. We intended to avoid blind spots in the algorithm; unusual and rare shapes could easily be “overlooked” during training, resulting in potentially poor algorithm performance on sites with unusual shapes. Instead, we created the site shape synthetically in software from a variety of 2D shape generators, such as:

- Dart Shape: Select random points on a canvas and form a convex hull around them
- Voronoi Shape: After seeding random seed points on a canvas, select for each of the seed points all points on the canvas that are closer to it than to any other seed point. Divides the canvas into convex polygons.
- Recursive Rectangular Subdivision: Creates an irregular shape by recursively splitting a rectangle and keeping a random subset of the resulting subrectangles.

Dataset 1

Dataset description

Dataset 1 is a predominantly tabular dataset where each row contains one particular site configuration, i.e., the site’s ID (referencing an external file that contains polygon coordinates), area, shape descriptors, width of potentially inserted streets, and the selected building typology, as well as the resulting SCR for this configuration. The dataset was constructed by feeding the sites into Hektar, which created subdivisions (parcels) and building footprints on the parcels, then calculating SCR from the overall site and footprint areas. Importantly, Hektar was deterministic at the time, meaning each configuration resulted in one precise SCR value.

This process was repeated for each building typology and a selection of street widths from very narrow to very wide, leading to a “full schedule” for each site: All SCR values for all possible configurations of a site were computed.

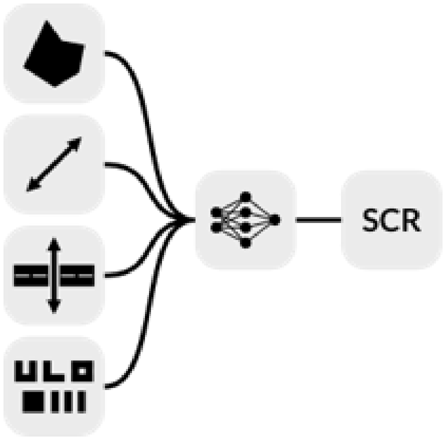

Image 15. *Default setup for SCR prediction tasks: SCR (output) is predicted based on site shape, size, street width, and building typology.*

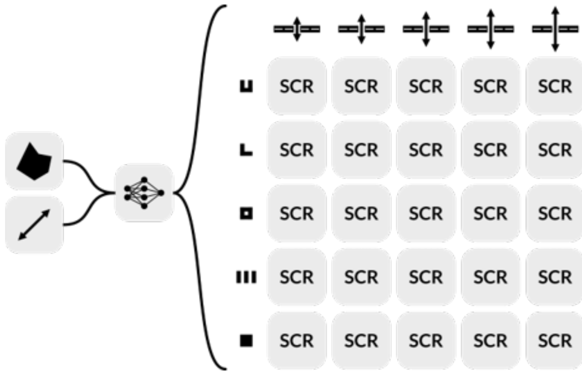

Image 16. *Alternate setup for SCR prediction tasks: SCR (output) is predicted based on site shape, size, street width, and building typology.*

Model setup

Image 15 shows the default setup for an SCR prediction task: The ML model is fed with a representation of a site's shape and size and other relevant variables (here: road width and building typology), and it outputs its estimate for the SCR as a single value (scalar).

However, the "full schedule" dataset allowed us to restructure the inputs and outputs of the model to a setup where only the site's shape and size are used as inputs. At the same time, SCRs for all possible configurations of street width and building typology are output simultaneously, as shown in image 16. The advantage of this setup for the user is that for a given site, only a single compute step is required to cover all possible configurations (varying street-widths and building typologies) at a low additional compute cost in the initial inference run, allowing for instant feedback when changing these settings.

Representations and model architectures

In dataset 1, we compared image-based vs. numerical representations of the shape. The image-based approach used a CNN model (Resnet 18), while the numerical approach used a small MLP with three hidden layers (see image 17).

Conclusion

The main takeaway from the study of dataset 1 is that the numerical representation of the shapes works surprisingly well, while being faster to compute and more straightforward in implementation than the image-based method. The image-based model could benefit from different representations of a site's scale/size, while the numerically based model could benefit from additional shape descriptors. While more aspects could have been evaluated on this dataset, the ongoing development of Hektar led to this dataset being outdated, and focus was shifted to the second dataset.

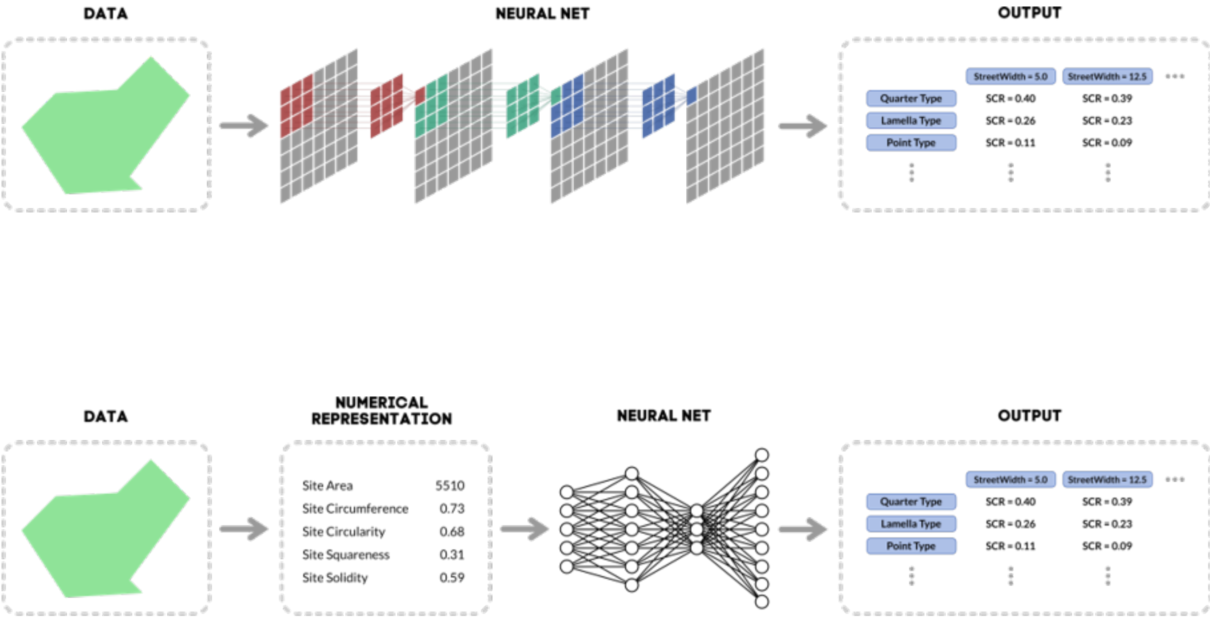

Image 17: *Model setups for dataset 1.*

Model	MSE	MAE	±0.1 SCR	±0.05 SCR	±0.02 SCR
CNN (image-based)	.0074	.0448	87.5%	74.9%	54.0%
MLP (numerical)	.0042	.0296	92.3%	83.2%	66.7%

Table 4: *Comparison of performance figures for CNN (image-based) and MLP (numerical). MSE: Mean Squared Error. MAE: Mean Absolute Error. ±X SCR: Percentage of predictions within a range of ±X from the actual value.*

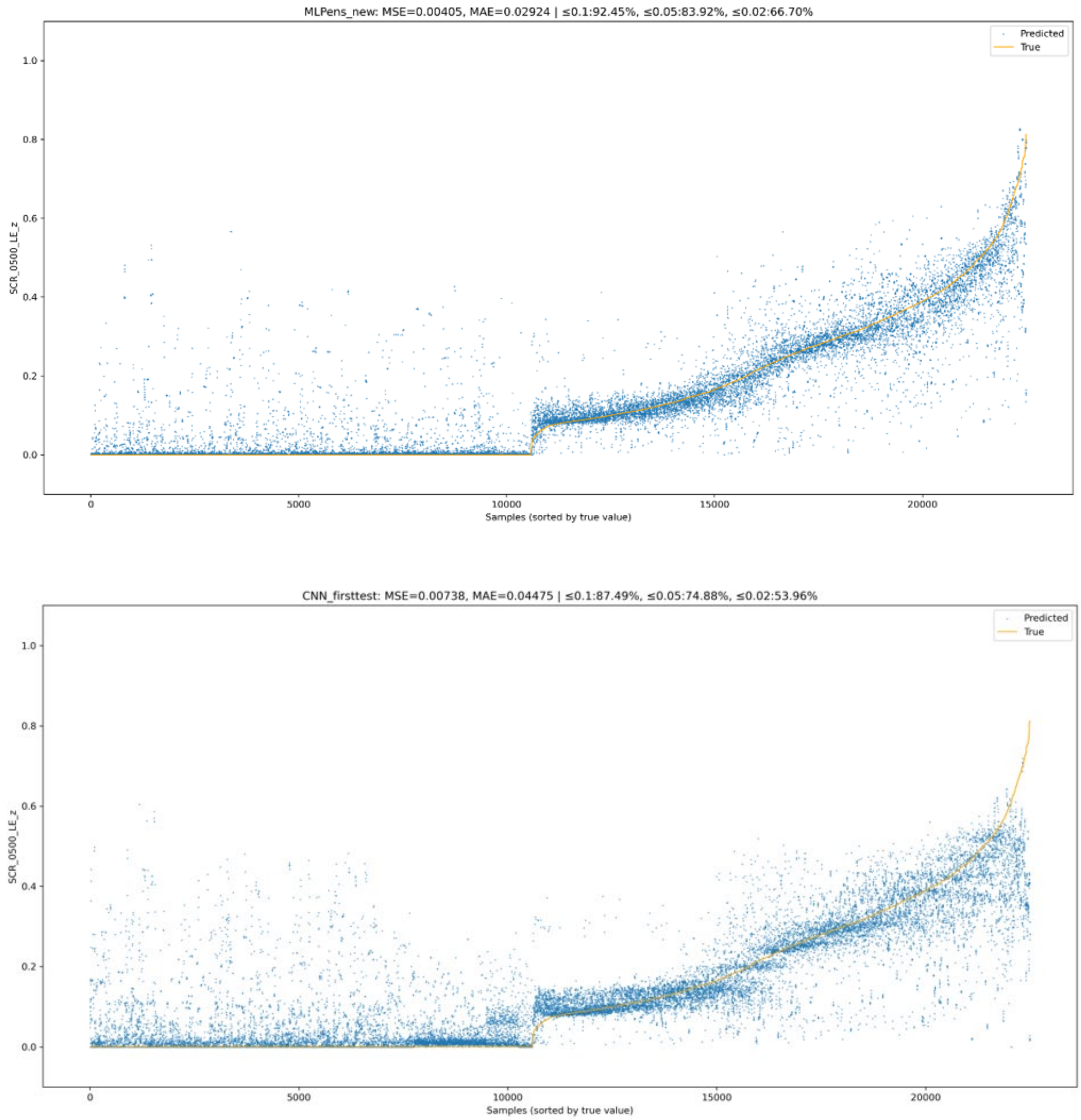

Image 18: *Visual representation of prediction accuracy between the two model types (values ordered by true SCR). The y-axis shows SCR, the yellow line shows true SCR, and the blue dots show predictions.*

Dataset 2

Changes to Hektar

The aforementioned changes to Hektar were an update to creating building footprints. Specifically, the available typologies were extended and restructured into six main typologies and dozens of sub-types per main typology, from which the algorithm takes a random selection at runtime. Therefore, the building footprint is no longer deterministic for a given typology, leading to a distribution of SCR values for each configuration instead of the individual numbers (scalars) from dataset 1. Moreover, the algorithm is no longer restricted to only one typology for the site. Still, it allows the user to select any combination of typologies for any given site (e.g., point houses, L- and O-shaped buildings can be combined and stand next to each other on the same site).

Dataset description

Dataset 2 is again a predominantly tabular dataset where each row contains one particular site configuration, i.e., the site's ID (referencing an external file that contains polygon coordinates), area, shape descriptors, width of potentially inserted streets, and the selected combination of building typologies, as well as the resulting SCR distribution for this configuration. The distributions were created by random sampling from repeated realisations of any given parcel configuration via Hektar: For each parcel and site configuration, one out of n building footprint areas was picked at random and added up to build a possible total building footprint area per site, and this was divided by the site's total area to receive a site-wide SCR value. This process was repeated 100x per site configuration to create a probability distribution over SCRs for any site configuration.

Due to the large number of possible combinations of typologies and street widths, and the need for repeated runs per configuration, we opted to move away from the "full schedule", running Hektar only for a selection of all possible configurations per site, and reduced the number of sites in the dataset to 668.

The distribution over SCR values was binned into 11 bins, with the first bin representing a precise "zero", and the remaining 10 bins equally spaced in the interval [0, 1].

Model setup

The model setup was changed from the one used in dataset 1. The approach of predicting all possible configurations' SCR was abandoned due to the low number of sites overall in dataset 2. Instead, the configuration setup was moved into the model's input variables, with the output being the 11-binned probability distribution over SCR values for any given configuration (see image 19).

Moreover, the study of dataset 2 focused on the numerical approach, and no image-based comparisons were made.

Results

To understand the usefulness of the model's predictions, we compared predicted versus real SCR distributions (see image 20). Given the small size of the dataset and the relatively low number of shape-descriptors, we rate the results as satisfying: On a random selection of test-set configurations, we observe good adherence to the target distribution, with the median matching in most cases. The predicted distributions, however, tend to be somewhat wider than the actual distributions.

Conclusion

A good match between actual and predicted distributions encourages and validates the practical viability of representing the site's shape via numerical descriptors to predict SCR values.

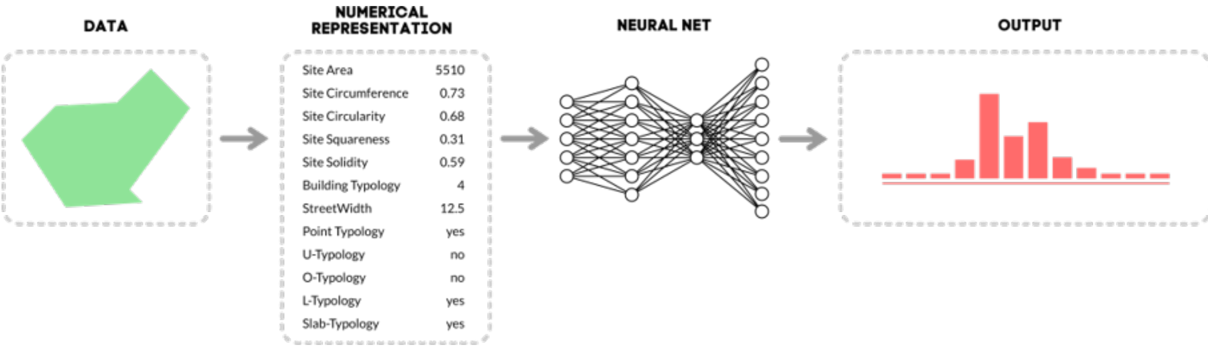

Image 19: *Model setups for dataset 2.*

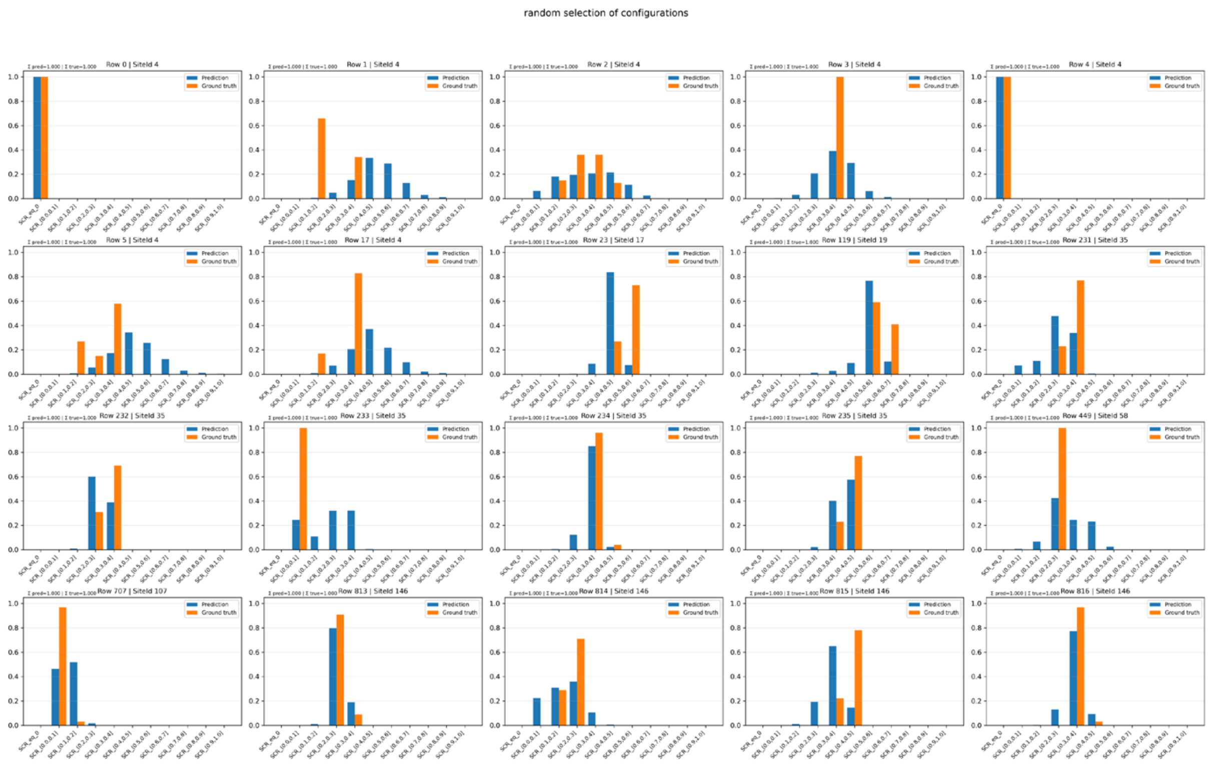

Image 20. *Resulting histograms of actual SCR values in orange and predicted values in blue.*

Image 21. *A presentation was held about the project at the Planning day in Iceland 2025.*

Communication

Communication activities aimed to bridge technical development with urban planning professionals' practical needs and perspectives. The team continuously tested and validated results in workshops with the reference group, using practical insights from urban planning as input for further development. We also communicated project progress through social media and professional forums to increase transparency and contribute to a broader discussion about the role of AI in urban development.

As part of the project's dissemination efforts, RISE presented the results at the National Planning Day Conference in Reykjavík, Iceland, in October 2025. The event gathered approximately 150 participants on-site and around 100 online attendees, representing planners, urban officers, and other professionals interested in using AI in planning practice.

Results

The project successfully implemented and validated two complementary computational approaches: a deterministic rule-based and data-driven machine learning model trained on synthetic datasets.

The deterministic model achieved very high precision. Nearly all test cases fell within ±5 % deviation between user-defined targets (FAR, SCR) and resulting outputs, confirming the system's reliability and reproducibility under defined constraints. The algorithm now performs pre-generation validation of all user inputs and flags infeasible configurations before geometry creation, ensuring that only geometrically and physically valid solutions are generated.

The machine learning study was conducted in two stages, corresponding to two generations of the Hektar system.

Stage 1 – Dataset 1

The first stage tested alternative ways to represent site geometry and typology. A convolutional neural network (CNN, Resnet 18) using image-based site representations was compared with a Multilayer Perceptron (MLP) using compact numerical descriptors such as area, solidity, rectangularity, circularity, and axis ratios. The numerical approach outperformed the image-based model across all metrics.

The MLP reached 92 % of predictions within ± 0.1 SCR and 83 % within ± 0.05 SCR, while being faster to compute and easier to interpret. This demonstrated that simple geometric descriptors are sufficient to capture site shape for the project.

Stage 2 – Dataset 2

An updated version of Hektar introduced stochastic building-footprint generation, producing a distribution of Site Coverage Ratio (SCR) values for each configuration rather than a single deterministic output. The corresponding dataset contained 668 sites with multiple combinations of typologies and street widths, each repeated 100 times to form probabilistic SCR distributions.

A refined numerical model was trained to predict these distributions in 11 bins over the interval [0–1]. The predicted and actual distributions aligned closely, with matching medians in most cases. The model slightly overestimated the spread, which is expected given the limited dataset size.

These results confirm the technical viability of using synthetic data to train AI models capable of rapidly estimating density outcomes in early-stage planning. The deterministic algorithm ensures interpretability, while the ML models add speed and the ability to represent uncertainty.

GitHub repository

https://github.com/ParametricSolutions/far-scr-prediction-model

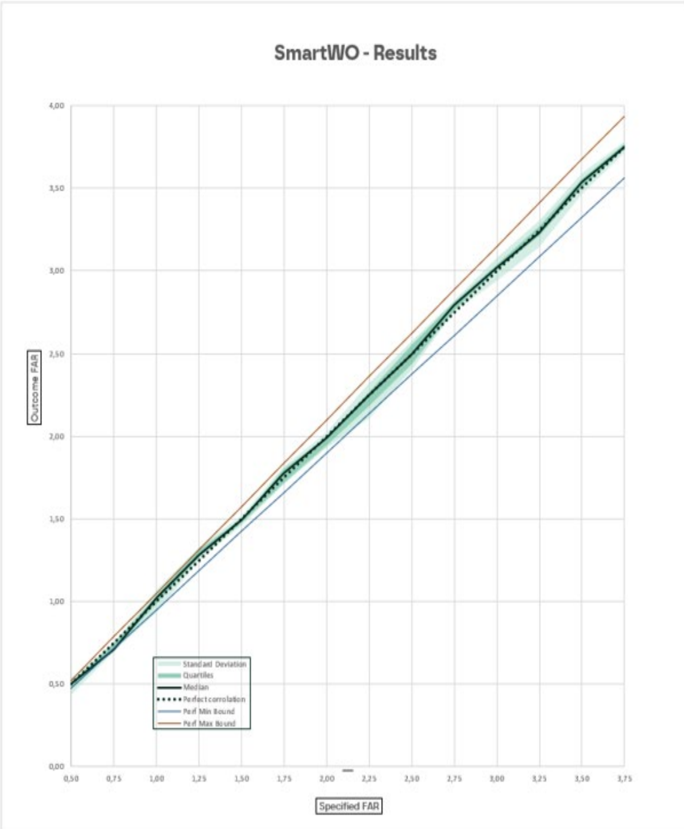

Image 22. *Deviation of FAR of desired and resulting numbers for the new Hektar algorithm.*

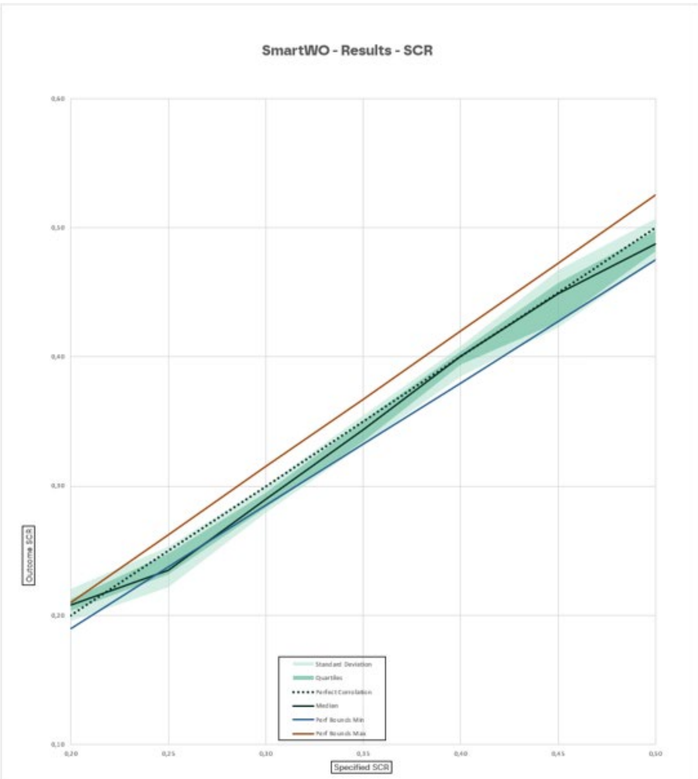

Image 23. *Deviation of SCR of desired and resulting numbers for the new Hektar algorithm.*

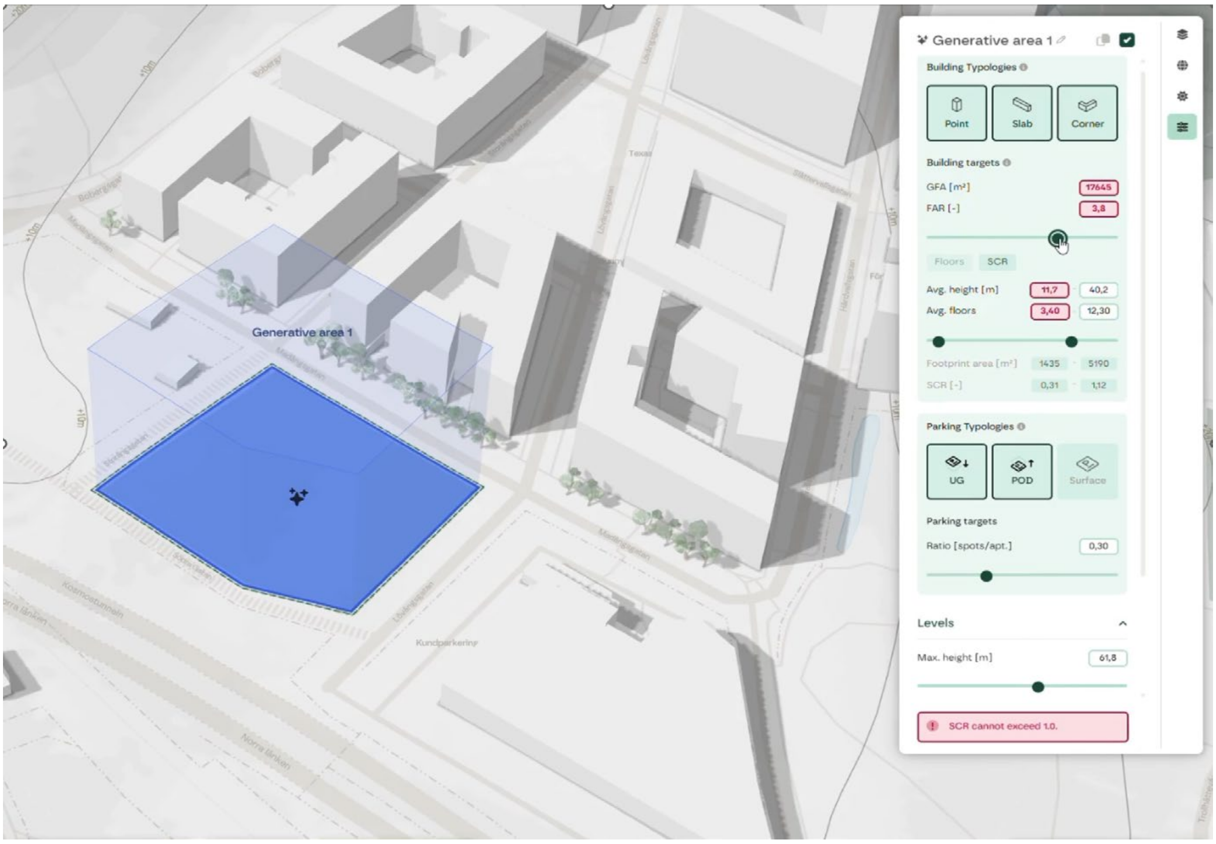

Image 24. *The improved deterministic algorithm now performs pre-generation validation of all user-defined parameters. It identifies combinations of density targets and geometric constraints that would result in infeasible configurations such as exceeding the maximum site coverage ratio (SCR). The system flags these conflicts before generation, preventing impossible "recipes" and ensuring that all generated outcomes remain within physically and geometrically valid limits.*

Discussion

The project shows that artificial intelligence can play an active, explainable, and practical role in early-stage urban development. Combining deterministic and machine-learning approaches provides complementary strengths: the rule-based model ensures consistency and transparency, while ML models allow for rapid feedback and probabilistic reasoning.

The two-stage ML development reflects the evolution of Hektar from a deterministic system to a stochastic, data-rich platform. This enables the project to explore both point-value and distribution-based prediction of density metrics. The results demonstrate that compact numerical shape descriptors are sufficient for predictive modelling, making the workflow efficient and easy to transfer across different contexts.

The probabilistic prediction introduced in the second stage is a significant methodological advancement. Instead of giving a single output, the model estimates a likelihood distribution over a feasible density outcome. This captures real-world uncertainty in early-stage planning and provides designers with a more realistic framework for decision-making.

The study also underscores the importance of high-quality synthetic data. While the datasets covered a broad range of shapes and typologies, including environmental and other parameters such as daylight, noise, and climate impact, would enhance the model's predictive ability and broad applicability. Validation against actual projects remains essential to ensure that models accurately reflect the diversity and regulatory nuances of real urban environments.

Integrating rule-based and probabilistic methods shifts the planning logic from form generates data to data generates form. This facilitates iterative scenario testing at a speed and scale beyond traditional workflows, supporting more informed, transparent, and evidence-based decisions during early design stages.

Although the environmental impact of AI systems is often debated, the models developed in this project are small-scale and lightweight compared to large generative AI models. Their energy use is modest, especially relative to the potential downstream benefits of more informed urban design decisions such as more efficient land use, less construction waste, and improved building performance. AI in urban planning covers a broad range, from high-performance image generation and simulation to specialised analytical tools like the one explored here. This variety of scales and purposes indicates that responsible and domain-specific AI can significantly reduce the impact of climate change on the built environment.

Next Steps

Further development of Hektar

Integrate the probabilistic prediction framework into the Hektar platform, allowing users to visualise uncertainty directly in the interface. Extend the model to additional sustainability parameters; daylight, noise, microclimate, carbon footprint, and material use to provide multi-criteria evaluation at the plot or district level.

Hybrid and real-world training data

Combine synthetic and real project datasets to improve model robustness and calibration. Benchmark predictive performance against verified development projects and regulatory constraints to ensure relevance for Swedish and international planning practice.

Algorithmic refinement

Explore ensemble and Bayesian architectures for probabilistic prediction, enhancing confidence estimation and model interpretability. Extend feature sets with topological and contextual descriptors such as street connectivity, accessibility, and surrounding land use.

Implementation and dissemination

Test the combined deterministic–ML workflow in ongoing municipal and developer collaborations. Document results through open publications and workshops. Position the tool for international uptake via partnerships and a scalable licence model.

Capacity building

Continue to develop competence in AI-supported planning through training sessions and practitioner courses, ensuring that urban planners and designers can apply these methods critically and responsibly in practice.

RISE Research Institutes of Sweden AB

Box 857, 501 15 BORÅS

Phone: 010-516 50 00

E-mail: info@ri.se, Internet: www.ri.se

RISE Report 2025:99

ISBN: 978-91-90036-88-4