Predicting Burglaries: A ML Approach for Low-Density Crime Areas

Luca Persia, Eduardas Lazebnyj, Andrea Günster, Damian Kozbur and Jérémy Decerle

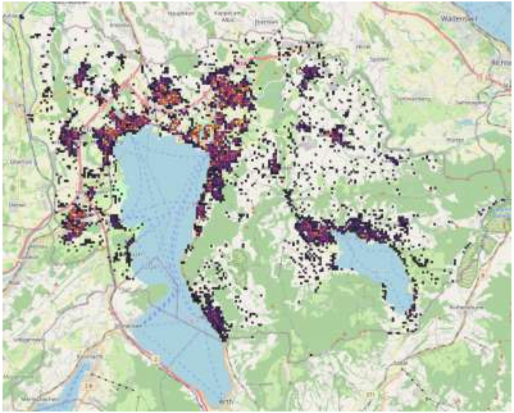
June 25, 2025

Predicting Residential Burglaries in Low-Crime Areas

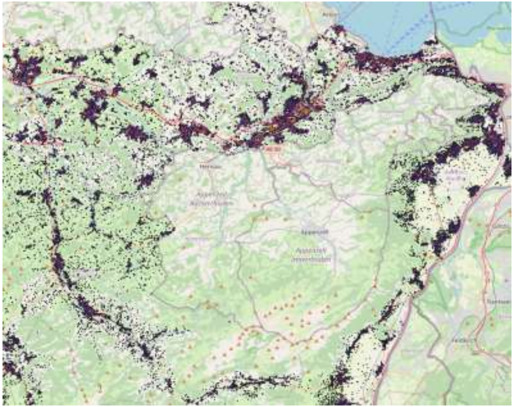
- **Low-risk areas:** Few or no recent crimes; traditional solutions are provided only for high-crime areas.
- **Goal - Why it matters:**
  - Early detection can prevent escalation.
  - Enables fair, proactive patrol planning.
  - Predictive policing enhance public safety.
- **Limitation of Near-Repeat Logic:**
  - Assumes crime clusters in time and space.
  - Ineffective where no recent crimes are observed.
- **Our Approach:** Daily ML-based predictions without needing recent local incidents.
- **Preliminary Results:**
  - High sensitivity in detecting risk areas.
  - False positives remain a challenge.

Data Sourcing & Engineering

Each grid represents a 100m $\times$ 100m area; lighter colors indicate higher residential density.



Zug Residential Housing (2012–2021)
Event Rate: **.0208%**, 2207 out of 10mln
Avg Yearly Event Rate: **.0220%**
Observed Near Repeat: **563**



St. Gallen Residential Housing (2012–2021)
Event Rate: **.0101%**, 7309 out of 72mln
Avg Yearly Event Rate: **.0108%**
Observed Near Repeat: **1319**

Vary in time and stable in space

- Temporal Factors
- Weather

Stable in time and vary in space

- Crime Pattern Theory
- Employement Statistics
- Population Statistics

Vary in time and vary in space

- Near and Canton Repeat

Feature Selection

Summary Statistics (ZG)

	mean	std	min	25%	50%	75%	max	skewness	kurtosis
month	6.628	3.540	1.0	4.0	7.0	10.0	12.0	-0.027	-1.240
year	2016.5	3.389	2012	2014	2016	2019	2022	-0.002	-1.152
popdens	34.769	41.599	3.0	5.0	15.0	52.0	339.0	1.793	3.468
swiss_popdens	25.275	29.613	0.0	4.0	11.0	38.0	242.0	1.843	3.954
nswiss_popdens	9.747	15.677	0.0	0.0	3.0	12.0	226.0	3.089	14.491
male_popdens	17.832	20.842	0.0	3.0	8.0	26.0	176.0	1.870	3.932
female_popdens	17.480	20.858	0.0	3.0	7.0	26.0	228.0	1.910	4.384
firmdens	8.046	16.353	4.0	4.0	4.0	5.0	405.0	8.320	100.968
empdens	40.874	124.697	4.0	4.0	5.0	19.0	3263.0	9.293	143.006
tavg	8.897	7.579	-13.4	2.6	8.8	15.4	27.9	0.003	-0.921
prcp	3.287	6.605	0.0	0.0	0.1	3.5	67.2	3.153	13.134
near_repeat	0.000	0.010	0.0	0.0	0.0	0.0	1.0	101.607	10321.938
nr_risk	0.075	0.264	0.0	0.0	0.0	0.0	1.0	3.224	8.396
cr_risk	0.0968	0.176	0.0	1.0	1.0	1.0	1.0	-5.309	26.185

Binary Classification

From Probabilities to Classes: A threshold (e.g., 0.5) is applied to the model output probability to classify each grid-day as **0 (no burglary)** or **1 (burglary)**.

Metrics:

	Pred 0	**Pred 1**
Act 0	TN	FP
Act 1	FN	TP

- **Precision:** $TP/(TP + FP)$ — How many predicted positives were correct
- **Recall (TPR):** $TP/(TP + FN)$ — Detection rate
- **FPR:** $FP/(FP + TN)$ — False alarm rate
- **AUC (Area Under ROC Curve):** Discrimination ability across thresholds (1 = perfect, 0.5 = random)

Evaluation: We use ROC curves to visualize the trade-off between TPR and FPR over all possible thresholds.

How Classifiers Learn

- Classifiers are trained to minimize a **loss function** on the training data (supervised learning).
- For binary classification, a common choice is the **binary cross-entropy loss** (a.k.a. log loss).
- Cross-entropy loss penalizes wrong or overconfident predictions:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i \log p_i + (1-y_i)\log(1-p_i)\Big),$$

  where $y_i \in \{0,1\}$ is the true label and p_i is the predicted probability $P(y_i = 1)$.
- The model adjusts its parameters to reduce L (using algorithms like gradient descent), thereby improving prediction accuracy on positives vs negatives.

Modeling Setup

Application:

- Regions: Cantons of **Zug** and **St. Gallen**
- Time range: **2012–2021** (daily spatial data at 100m $\times$ 100m resolution)

Training and Evaluation Procedure:

- **Resampling:** Random Undersampling is applied to handle the strong class imbalance
- **Cross-Validation:** 5-fold cross-validation on data from 2012–2019 for hyperparameter tuning
- **Test Set:** 2019–2021 held out for final evaluation

Feature Subset:

- **Continuous:** popdens, swiss_pop, nonswiss_pop, male_pop, female_pop, businesses, empldens, tavg, prcp
- **Categorical:** month, year
- **Binary:** near-repeat and cantonal-repeat features

Baseline Approach

Procedure:

- All models are trained using **5-fold cross-validation** on the training set (2012–2019)
- Randomly sample $\mathcal{D}_b^-$ from the majority class (non-burglaries), with $|\mathcal{D}_b^-| = |\mathcal{D}^+|$

Models Evaluated:

- Logistic Regression $->$ interpretable linear baseline
- Random Forest and LightGBM $->$ non-linearities and complex interactions

Ensemble Approach

Procedure (Kadar et al., 2019):

1. For $b = 1, \ldots, B$ ensemble iterations:
   - Randomly sample $\mathcal{D}_b^-$ from the majority class (non-burglaries), with $|\mathcal{D}_b^-| = |\mathcal{D}^+|$
   - Construct balanced training set $\mathcal{D}_b = \mathcal{D}^+ \cup \mathcal{D}_b^-$
   - Train and 5-fold cross-validation of the base learners f_b on $\mathcal{D}_b$
2. Aggregate predictions over all B models:

$$\hat{p}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x)$$

Cost-Sensitive Learning Approach

Procedure (Focal Loss, Lin et al., 2017):

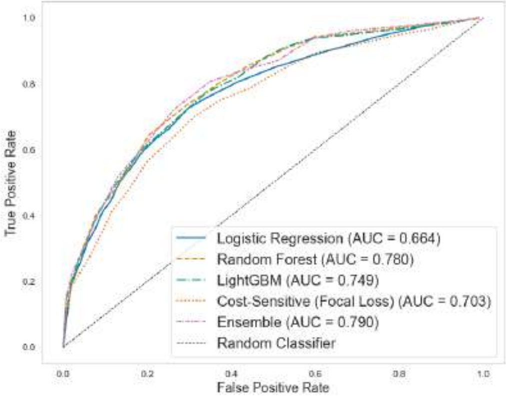
$$\mathcal{L}_{\text{Focal}} = -\alpha \cdot (1-p)^{\gamma} \log(p)$$

where $p = P(y = 1 \mid x)$ is the predicted probability, $y \in \{0, 1\}$ is the true label, $\alpha \in (0, 1)$ introduces class weighting (with $\alpha < 0.5$ emphasizing the minority class), and $\gamma > 0$ down-weights easy examples, focusing learning misclassified instances.

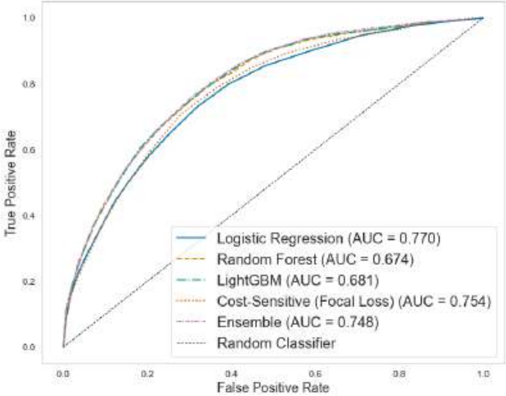
Note: Focal Loss is inherently class-balanced via α, and reduces to standard cross-entropy when $\gamma = 0$, i.e., $\mathcal{L}_{\text{Focal}} = \mathcal{L}_{\text{CE}}$.

Implementation: This loss function is applied only to the LightGBM model, as it allows direct customization of the objective function and supports gradient-level loss tuning.

Results: Zug vs St. Gallen



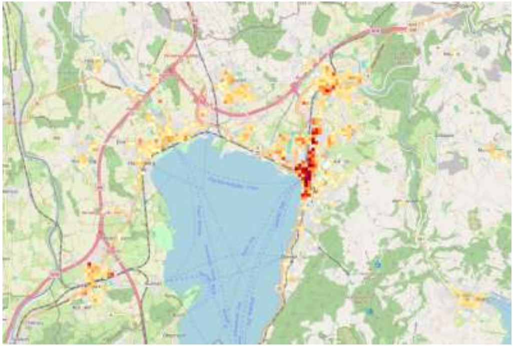
Zug Canton - ROC Curves



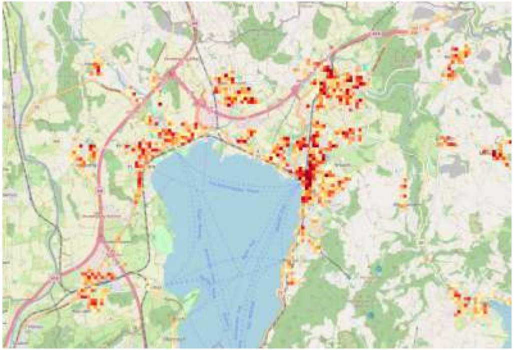
St. Gallen Canton - ROC Curves

Results: Hotspots

Zug — 12 January 2019



Ensemble



Cost-Sensitive

Color gradient: Only predictions above 0.5 are visualized. Darker intensity indicates increasing probability, ranging from 0.5 (light red) to 1.0 (deep red).

Results: Hotspots

St. Gallen — 12 January 2019

Ensemble

Cost-Sensitive

Color gradient: Only predictions above 0.5 are visualized. Darker intensity indicates increasing probability, ranging from 0.5 (light red) to 1.0 (deep red).

Outlook

Extension of the Current Approach

- **Scaling to Additional Cantons** Data access for Aargau and Zurich is expected in the coming months, allowing testing of model generalizability across diverse regions.
- **Methodological Considerations**
  - Design of **tailor-made loss functions** aligned with operational needs of police units
  - Implementation of **adaptive thresholds** based on local population density or urban structure
- **Computational Constraints** Due to memory and runtime limitations, only a subset of features was used in training. Access to more scalable infrastructure is needed to fully exploit the available data and model complexity.

Future Research I

Expanding the methodology toward spatio-temporal point processes

GIS Building-Level Granularity

Zug City - Buildings Map

From Grids to Point Processes

- Traditional ML models operate on discrete spatial grids, assuming independence across time and space.
- **Point processes** provide a more natural framework for modeling the timing and location of crime events, treating them as realizations in continuous space and time.
- We focus on **Hawkes processes** — self-exciting models where each event increases the short-term likelihood of future events (e.g., aftershocks following an earthquake in seismology):

$$\lambda(t \mid \mathcal{H}_t) = \mu + \sum_{t_i < t} g(t - t_i)$$

where $\lambda(t)$ is the event intensity at time t, μ is the background rate, and $g(t - t_i)$ quantifies how past events raise future risk.

Temporal Hawkes Process — Zug City I

Log-likelihood function for a univariate Hawkes process with kernel $g(t) = \alpha e^{-\beta t}$, assuming $T = t_n$:

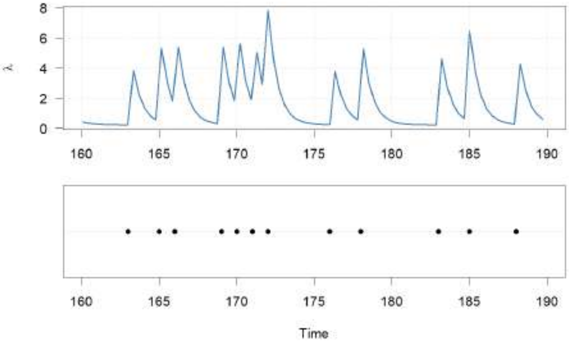
$$\log L(t_1, \ldots, t_n \mid \theta) = -\mu t_n + \sum_{i=1}^{n} \frac{\alpha}{\beta} \left(e^{-\beta(t_n - t_i)} - 1 \right) + \sum_{i=1}^{n} \log \left(\mu + \alpha A(i) \right)$$

with $\theta = (\alpha, \beta)$, $A(i) = \sum_{t_j < t_i} e^{-\beta(t_i - t_j)}$ and $A(1) = 0$. (Ozaki, 1979)

Parameter	**Estimate**	**Std. Error**	**p-value**
α (excitation)	3.82	0.88	< 0.001 ***
β (decay rate)	1.62	0.62	0.009 **

MLE via BFGS. Results for a univariate Hawkes process (Zug City, 2016), with fixed $\mu =$ average baseline daily intensity $= 0.22$.

Temporal Hawkes Process — Zug City II



Zug City (2016): estimated temporal intensity. Burglary intensity exhibits self-excitation — a single event temporarily raises the risk of follow-up events, indicating repeat-victimazion.

Next steps: (i) extend to spatio-temporal Hawkes processes (Reinhart, 2018), modeling $\lambda(s, t \mid \mathcal{H}_t)$, (ii) extend to marked Hawkes process (e.g., different crime types), include contextual covariates to enrich background rate μ

Future Research II

Expanding the current database

Weather and Light Conditions

Weather for Zug Canton (real)									
City	Lat	Lon	Date	Sunrise	Sunset	Moonrise	Moonset	Illum.	Condition
Zug	47.17	8.52	2025-04-08	06:51	20:05	15:29	05:23	78	Overcast
Zug	47.17	8.52	2025-04-07	06:53	20:04	14:17	05:01	69	Sunny
Zug	47.17	8.52	2025-04-06	06:55	20:03	13:03	04:34	59	PartlyCloudy

MaxT	MinT	AvgT	Wind	Precip	Snow	Vis	Hum	UV	Rain%	Snow%	WillRain
14.2	-2.5	5.5	9.0	0.00	0.0	10.0	64	3	0	0	0
12.8	-5.1	3.5	10.1	0.00	0.0	10.0	60	4	0	0	0
12.2	1.0	6.9	15.5	0.00	0.0	10.0	71	4	0	0	0

Road Network Topology (OSM)

Road Segment Data – Zug (real)									
Start	End	Length (m)	Type	Name	Junction	Speed	Bridge	Tunnel	Oneway
x_1	y_1	214.68	p	...	none	50	no	no	False
x_2	y_2	52.00	s	...	none	60	no	no	False
x_3	y_3	664.14	p	...	none	80/50	no	no	False
x_4	y_4	17.99	s	...	roundabout	unknown	no	no	True

Live Traffic

Live Traffic Segment – Zug (real)					
FRC	Curr. Speed (km/h)	FreeFlow Speed	Travel Time (s)	FreeFlow Time	Confidence
FRC4	23	35	218	143	1.00

- `FRC` = Functional Road Class (0 = motorway, 5 = local roads)
- **Congestion signal:** Curr. Speed $\ll$ FreeFlow Speed $\Rightarrow$ high likelihood of human presence
- Combined with OSM to create a dynamic exposure map

Crowd Dynamics

- **Data source:** anonymized, aggregated telecom mobility from Sunrise
- **Spatial resolution:** 50–100m in urban areas
- **Temporal resolution:** 15 minutes to daily
- **Demographics:** age, gender, nationality, inferred home/work location

Daily Mobile Device (hypothetical sample)

Date	ID	Total Devices	Home Share	Tourists	Median Age	Male%	Foreign%
2023-07-01	..	412	0.74	15	39.2	0.51	0.28
2023-07-02	..	398	0.76	18	40.1	0.50	0.29
2023-07-03	..	423	0.73	11	38.8	0.52	0.27

Results & Discussion

Machine Learning models

- Relatively easy to implement, but with high trade-off between hit rate and FPR
- Limited spatial-temporal resolution and contextual features
- There are potential improvements: (i) on spatial moving threshold and ad-hoc loss function, (ii) on testing generalizability with **Zürich and Aargau** data

Future Research I: Spatio-Temporal Point Processes

- Capture self-exciting dynamics using Hawkes processes
- Model crime as a continuous process in time and space
- Generalizable to different crimes (marked point process) and fine-grained hotspot detection

Future Research II: Dataset Enrichment

- Integrate building-level data (weather, sunrise, traffic and streets information)

Dissemination & Conferences

- **MSc Management and Law International Conference** — Lucerne, 16–17 Jan 2025
  Talk: "Quantifying Illegal Activity: Predicting Burglaries in Switzerland"
- **Conference UZH, Police Bureaus of Zurich, St. Gallen, Zug, Aargau** - University of Zurich, 7 Mar 2025
- **CEPOL Research & Science Conference** — Rome, 25-27 Mar 2025
- **Swiss Society of Economics and Statistics (SSES) Congress 2025**
  KOF, ETH Zurich, 26–27 Jun 2025
  Accepted Paper: "Modeling Crime Dynamics: Predicting Burglaries in Canton Zug"

Submission and Conclusion of the QIA Innosuisse Project

Prediction - Crime

Machine Learning - myABI Application I

- Easy to implement police intuition
  - Customized Integration for Officers to set radius, time, number of hot spots
  - customized loss function
- City Police Zürich and Cantonal Police Aargau

Spatio-Temporal Point Processes - myABI Application II

- Captures self-exciting dynamics
- Crime as a continuous process in time and space

Dataset Enrichment

- Integrate myABI Customers and use all CH crime data
- Expand to additional areas of crime which is predictable in time and space

Estimating Magnitude - Crime CH

Flexible Method (Available ✓)

- Dark Rate of **Cybercrime in CH** (Fedpol) ✓
- Dark Rate of **Product Piracy** (CH Customs)
- Dark Rate of Domestic Violence (City Police Zürich Prevention)
- Dark Rate of Different Corporate Crime (Cantonal Police Zug)

Dark Rate - myABI Application I

- Dark Rate of Cybercrime for firms and industries ✓
- Dark Rate of other Offenses

Estimating Magnitude - Crime EU

Flexible Method (Available ✓)

- Dark Rate of **Cybercrime in EU** (EUROPOL)
- Dark Rate of **Illegal Price-Fixing Cartels** (European Commission and Swiss Competition Commission) ✓
- Dark Rate of **Tax Evasion or Public Expenditures** (Guardia di Finanza)
- Dark Rate of **Tax Evasion** (Swedish Tax Authority)
- Dark Rate of **Insider Trading in Norway** (Per Östberg (UZH)) ✓
- Dark Rate of **Insider Trading in Norway** (CH FINMA)
- Dark Rate of **Insider Trading in USA** (US Stock Exchange Commission (SEC) (Bloomberg & Refinitiv))
- Dark Rate of **Product Piracy** (CH Customs)

Scientific Advancements

Methods

- Dark Rate of **Collusion and Cybercrime** - UZH PhD Nicole Bellert (Submission Jan 2026) (with Günster (ZHAW) and Sven Seuken (UZH))
- Self-exciting point process estimation of **Crime** and in Finance - USI PhD Luca Persia (with Günster (ZHAW) and Deborah Sulem (USI))
- Predicting Crime with Neural Networks and Transformers - USI PhD Eddy Lazebnyj (with Günster (ZHAW) and Ernst Wit (USI))

Applications

- EC and CH Crime
- Finance
- Economics
- Data Science and Stats