Machine Learning for Identifying Potential Photovoltaic Installations on Parking Areas

Frederick Kistner*[a] and Sina Keller*[b]

Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), Engler Street 7, D-76131 Karlsruhe, Germany

Keywords: Artificial Intelligence, Renewable Energy, Classification, Urban Areas, Sustainable Development, PV Installation.

Abstract: Integrating renewable energy systems into urban areas is crucial for sustainable development. This study assesses the potential for installing photovoltaic (PV) systems in parking areas, focusing on a case study region in Hesse, Germany. A machine learning approach is developed to classify parking lots larger than $900\,\mathrm{m}^2$ into suitable and unsuitable categories. The input data includes OpenStreetMap (OSM), the Authoritative Topographic-Cartographic Information System (ATKIS), and high-resolution geospatial datasets. A reference dataset for the two classification categories is created. Multiple input features are generated, and their significance for the classification task is evaluated. Additionally, several shallow machine learning models are implemented and assessed. The XGBoost model demonstrates the highest accuracy at 99 % and is used to classify 10,894 parking areas throughout Hesse. Key suitability features include the Normalized Difference Vegetation Index (NDVI), surface sealing ratios, and vegetation height. The results indicate that approximately $21.8\,\mathrm{km}^2$ of the parking area is suitable for PV installations, requiring minimal ecological intervention. The methodological approach is scalable for application in other regions, and validation in Frankfurt am Main confirms a strong correlation with solar radiation levels. This study provides a data-driven framework for optimizing urban energy systems and supporting sustainability initiatives.

1 INTRODUCTION

Renewable energy technologies are essential to mitigate climate change and ensure energy security (IPCC, 2022). Among these, photovoltaic (PV) systems stand out due to their scalability and adaptability to diverse environments (Santamouris, 2020). The scenarios analyzed by Fraunhofer ISE (Wirth, 2023) estimate that achieving climate neutrality in the German energy sector will require PV capacities ranging from 215 GW to 500 GW, depending on efficiency improvements, public acceptance, and energy system dynamics. In recent decades, solar power costs have decreased significantly, making large-scale plants highly competitive with fossil fuels (Wirth, 2023).

When focusing on urban regions, areas functional for PV installations are lacking. Parking lots are generally seen as underutilized areas or *stranded assets*, occupying substantial urban land that is mainly nonproductive (Krishnan et al., 2017). Transforming these areas with solar canopies offers a dual-purpose solution that encourages using renewable energy without requiring additional land resources (Ivanova et al., 2020). Research indicates that PV installations in these areas can substantially boost local energy production and contribute to sustainable urban development (Maier et al., 2024; Marneni et al., 2021; Krishnan et al., 2017).

An important question is how to identify parking lots suitable for installing PV. Current evaluations of parking lot solar installations often depend on generalized assumptions and lack detailed, high-resolution data. This lack can lead to overlooking local constraints and complicating feasibility analyses for PV installations in different parking lot environments. It could also eliminate parking lots, which could be valuable for solar installations.

The proposed study addresses these emerging aspects. While previous studies have explored the feasibility of solar energy installations in urban areas, no existing approach combines high-resolution geospatial data with machine learning (ML)–based classification to identify PV potential in parking areas. This study introduces a novel, automated ML pipeline

[a] https://orcid.org/0009-0002-2601-9538
[b] https://orcid.org/0000-0002-7710-5316
*These authors contributed equally to this work.

Kistner, F. and Keller, S.
Machine Learning for Identifying Potential Photovoltaic Installations on Parking Areas.
DOI: 10.5220/0013476300003935

In *Proceedings of the 11th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2025)*, pages 244-252
ISBN: 978-989-758-741-2; ISSN: 2184-500X

that leverages OpenStreetMap (OSM), the Authoritative Topographic-Cartographic Information System (ATKIS), and high-resolution geospatial datasets to systematically classify parking areas based on suitability for PV installation. We apply advanced preprocessing and feature engineering techniques to capture site-specific details. Robust classification algorithms, including the XGBoost, help differentiate suitable parking lots from unsuitable ones for PV canopy installations. This methodological approach goes beyond static, threshold-based filters by using ML to manage complex, multi-dimensional data.

The main contributions of this study are summarized:

1. **Integration of Heterogeneous Spatial Data:** We unify datasets from OSM, ATKIS, and high-resolution satellite imagery, including normalized difference vegetation index (NDVI), to create a comprehensive feature space containing descriptors for each parking lot.
2. **Automated Preprocessing and Feature Engineering:** We apply automated spatial analysis techniques, such as buffer-based processing, high-resolution land cover segmentation, and geometric feature extraction, to enhance data quality and improve model performance.
3. **Scalable Classification:** We conduct comparisons of supervised machine learning models, which are trained and evaluated on over $1{,}000$ labeled parking lot examples.
4. **Generic, Transferable Framework:** The methodology is designed to be replicable, allowing for broad applicability in large-scale assessments of PV potential.
5. **Independence from Solar Radiation Models:** A case study illustrates that the classification framework does not rely on solar radiation models, making it especially advantageous for regions without such data.
6. **Actionable Insights:** The analysis provides detailed insights into PV potential by considering sealing rates, vegetation heights, and other parking lot-specific attributes rather than relying solely on simplified metrics such as percentage area reduction.

2 RELATED WORK

This section briefly introduces related work on PV systems in parking lots and their feasibility. First, we present the facts and benefits of installing PV systems in parking lots. Second, we summarize the economic and structural aspects. Lastly, we look at existing spatial approaches to identify and classify suitable parking lots for PV.

PV Systems in Parking Lots: Generation and Consumption of Energy. In general, PV systems can generate power close to where it is consumed, thereby minimizing the need for extensive transmission lines. This setup becomes particularly effective when paired with distributed battery storage and other energy converters (Wirth, 2023). Parking lots have become ideal sites for installing decentralized photovoltaic systems. They allow battery storage integration, which helps ease grid congestion during peak production times and reduces the need for long-distance transmission. Consequently, many studies prioritize parking lots for their land use efficiency (Maier et al., 2024; Wirth, 2023; Solar Cluster BW, 2022; Figueiredo et al., 2017).

PV Systems in Parking Lots: Economical and Structural Aspects. While rooftop PV solutions are well-established, canopy structures in parking lots involve additional costs for support and retrofitting (Figueiredo et al., 2017). However, future cost reductions and increased incentives for on-site usage, especially when combined with the charging of electric vehicles, are expected to improve the economic viability of such installations (Solar Cluster BW, 2022; Maier et al., 2024). Although structural investments can be approximately 50 % higher than those for standard rooftop systems, on-site self-consumption and premium electricity pricing can offset these initial costs (Krishnan et al., 2017; Maier et al., 2024). Indirect benefits include vehicle protection, extended pavement life, and a visible demonstration of commitment to climate action, strengthening public and corporate perception (Solar Cluster BW, 2022).

Challenges in Identifying Suitable Parking Lots for PV Installation. Generalized assumptions are often made when focusing on existing studies and approaches to identifying potential parking lots for PV installation. These include, for example, fixed area coverage factors or capacity densities and a lack of high-resolution data on shading, vegetation, or land use constraints (Ludwig et al., 2024; Maier et al., 2024; Krishnan et al., 2017). Besides, parking lots vary widely in ownership structures, functional zones, and site conditions, complicating feasibility analyses for PV installations (Solar Cluster BW, 2022; Maier et al., 2024). For example, in Germany, several federal states mandate parking PV systems for newly constructed lots exceeding specific size thresholds, influencing scalability and business models, as shown by (Maier et al., 2024; Ludwig et al., 2024). In addition, automated large-scale methods for identifying and classifying suitable existing parking lots remain limited. Consequently, local limitations are often overlooked, leading to inaccurate potential assessments (Wirth, 2023).

3 DATASETS

Both input and reference data are essential for creating and evaluating ML approaches to identify potential parking lots for PV installations. We detail the reference and input data in Sections 3.2 and 3.3.

Since ensuring the feasibility and generalization of ML models is crucial, the implementation needs to be performed on a necessarily large actual dataset. Therefore, we train and test the ML models on a representative input-reference dataset that encompasses the parking areas of the German federal state of Hesse. In the subsequent Section 3.1, we give a brief overview of the geographic background of Hesse.

To further validate the ML approach, we investigate solar radiation information of selected parking lots. Section 3.4 summarizes these data.

3.1 Study Region

Hesse (German: Hessen) is a federal state bordered by six other states in west-central Germany. Its capital is Wiesbaden, and its largest city is Frankfurt am Main, a significant financial hub. Covering around 21.114 km^2, Hesse has a population of over six million residents. The landscape features hilly terrain and extensive forests, with about 42 % of its land area covered by woodlands. The Rhine River forms the southwestern border, contributing to its varied topography.

To promote sustainability, the Hessian Ministry of Economics, Energy, Transport, Housing and Rural Regions (German: Hessisches Ministerium für Wirtschaft, Energie, Verkehr, Wohnen und ländlichen Raum (HMWEVW)) claims that new parking lots with over 50 spaces must now include solar panels, with an expected 100 new solar-equipped lots each year (HMWEVW, 2023).

Hesse's central location, diverse landscape, and innovative energy policies make it an ideal region for developing our proposed ML models, which focus on potential PV installations in parking areas.

3.2 Generation of Reference Parking Lot Data

ATKIS and OSM are the primary data sources for parking lot polygons. One advantage is that these sources are freely available, ensuring the generalization opportunities of the developed approach.

OSM parking areas are identified using tags such as `parking`, `capacity`, `access`, `surface`, `rooftop`. To meet the requirements according to the Hessian Ministry of Economics, Energy, Transport, Housing and Rural Regions (HMWEVW, 2023), the minimum area of the parking lots needs to cover 900 m^2, approximately 50 spaces. Currently, relevant parking lots that meet the same size criteria are extracted from ATKIS. The OSM and ATKIS parking lots have been merged, and duplicates have been removed. This first step leads to $11,281$ parking objects covering a total area of 35.88 km^2.

For reference data generation, selected parking lots are manually labeled and divided into two categories for PV installations: *suitable* and *unsuitable*. This labeling process is based on high-resolution satellite images for $1,002$ parking lots, most of which are randomly chosen. The reference dataset comprises 775 *suitable* and 227 *unsuitable* parking lots. Table 1 visualizes examples for those two classes. Unsuitable parking lots constitute the minority class and are significantly underrepresented in the randomly selected fraction of the dataset, which posed challenges for solving the classification task with ML models. Therefore, additional examples of the minority class were systematically added to the dataset, enabling more robust model training. The resulting reference dataset comprised 22.6 % unsuitable and 77.4 % suitable parking lots stored as a geographic layer.

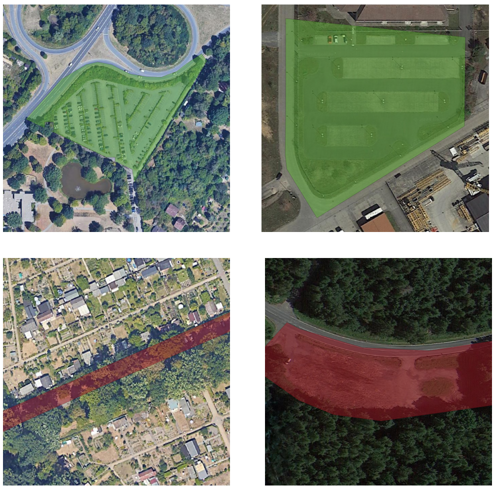
3.3 Heterogeneous Input Data Sources

In addition to the generated reference data (see Section 3.2), different input features must be extracted from various data sources.

Since we aim to analyze the potential of installation sites in Hesse, we are built upon a comprehensive dataset comprising various geometrical and environmental factors.

The primary data source includes $11,574$ **parking lot objects** described in Section 3.2. These parking lots serve as a foundational basis for identifying suitable locations for PV installations. Almost 9 % of these parking lots have labels. All parking lots regarding the **slope** and orientation are investigated. Excluded areas are characterized, for example,

Table 1: Exemplary parking lots manually labeled and classified as *suitable* (first row), and *unsuitable* (second row).



by a north slope $>5°$ or a steep slope $>30°$ (see Section 4.1).

Additionally, a raster file at a resolution of 0.2 m × 0.2 m has been generated, providing class predictions at the pixel level from a deep learning segmentation model.[1] Based on these results, we can calculate ratios within the parking lots and surrounding locations and extract information about the surface texture. In total, the land cover layer consists of eleven classes, such as fully `sealed`, `partially sealed`, `tall vegetation`, or `low vegetation`. A raster dataset of **NDVI** with a spatial resolution of 10 m × 10 m is employed to assess vegetation density and health. NDVI is calculated as the mean during the summer months between 2018 to 2023. Lastly, we include the total **green volume** with a spatial resolution of 100 m × 100 m calculated with an NDVI threshold and a normalized digital surface model.

Based on these input data sources, we generated several input features, as described in Section 4.

3.4 Additional Validation Data

To enhance and validate our proposed approach, we utilize average annual solar radiation data from 193 selected parking lots within the city district of Frankfurt am Main (approximately 20 % of Frankfurt's parking lot area). These data have been obtained by manually outlining the polygon shapes of the selected areas based on high-resolution solar radiation information provided by the Hessian solar register (German: Solarkataster Hessen) (Landes Energie Agentur Hessen, 2025). This online tool is designed to assess the suitability of rooftops and open spaces for PV installations, considering factors such as solar radiation, shading, and orientation (Landes Energie Agentur Hessen, 2025).

[1]This deep learning segmentation model was a result of the extended research within the project.

4 METHODS

This section outlines an automated pipeline for preparing, analyzing, and modeling parking lots to assess the suitability of PV installations.

4.1 Preprocessing

Nine buffer zones are established around 10.894 potential parking lots at distances of 0 m, 1 m, 5 m, 7 m, 10 m, 15 m and 20 m to facilitate the parking lots themselves (0 m) and their surrounding environment analysis. These buffers are dissolved with different input sources such as land cover, NDVI, and green volume (see Sections 3.3 and 4.3), and intermediate layers are generated for each buffer distance.

4.2 Splitting in Training and Test Dataset

Based on the labeled reference data, we split the dataset using stratified sampling with a ratio of 80 : 20 into a training and test set using scikit-learn (Pedregosa et al., 2011). In addition, we apply a fixed random seed to ensure reproducibility. The imbalance of the dataset is the main reason for the stratification. In pre-testing, we have systematically added labeled data points from the minority class to the dataset. This manual extension effectively implements an upsampling approach with actual data, addressing the dataset imbalance (see Section 3.2) (More, 2016).

The test set consists of parking lots the models have never encountered. It is used exclusively for final evaluation and has never been part of the training phase. As outlined below, the 80 % portion designated for training is split into three cross-validation folds for hyperparameter optimization.

Figure 1 shows the distribution of unsuitable and suitable parking lots in the training and test sets.

4.3 Feature Extraction

Based on different input data sources, we extract additional features.

To extract the features, the calculated buffer zones overlap with the thematic layer NDVI, green volume,

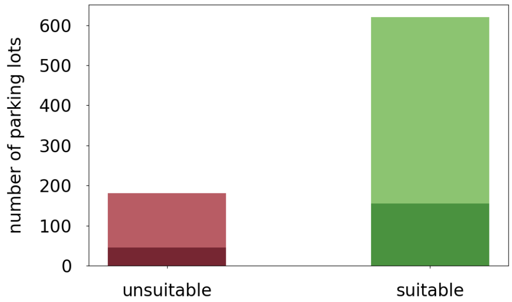


Figure 1: Class distribution of the training and test set. The unsuitable parking lots are visualized in red, while the suitable parking lots are green. The respective darker colors represent the test subset.

and land cover information. Therefore, we calculate **five key geo-statistics features**: minimum, maximum, mean, median, and sum for each buffer area and the NDVI and green volume.

Regarding the high-resolution land cover, we extract **three geostatistical features per buffer zone**: count, area, and proportion for the eleven land cover classes. The count is given in pixel numbers, while proportion represents the ratio between the occurring land cover classes.

With every overlapping and calculation, intermediate layers are created and merged into one layer via spatial join, and new attributes are assigned for each buffer distance and topic.

In a subsequent feature extraction step, we compute additional geometry-related features for each parking lot object. These include bounding box dimensions, e.g., `lengthwidth`, `elongation`, `perimeter`, `compactness`, `convexity ratio`, `centroid coordinates`, and `solidity`. The **bounding-box area-to-lot area ratio** captures the object's fit within its bounding box, while the shape area and perimeter are updated accordingly. Furthermore, **three additional class ratio features** are generated, relating sealed (fully or partially sealed) pixel counts to vegetation counts (low, medium, or tall).

Finally, all existing NDVI and green volume features are normalized by each lot's area, producing respective variants divided by shape area and further enhancing the comparability of metrics across differently sized parking lots.

After applying correlation analysis, we discard highly correlated features with a correlation coefficient >95 %, resulting in 70 final input features. Additionally, we have tested dimension reduction techniques without considerable impact on the models' performance.

In the final step, we remove 77 parking lots with missing values due to district boundaries exceeding limits. We then combine the input features with the suitability class labels of the reference dataset (1,002 objects), which contain information about whether the dataset will be used for training or testing. The remaining 9,892 parking lots are not labeled and will be applied in the final model.

Within the Frankfurt city district, we combine the solar radiation data (see Section 3.4) with the labeled parking lots to evaluate the model's predicted suitability classes and our manually labeled reference class later.

4.4 Model Development and Optimization

Several ML methods exist for supervised learning. We study selected state-of-the-art **shallow learning ML** approaches to solve the classification task: Random Forest (Breiman, 2001), XGBoost (Chen and Guestrin, 2016), Extra Trees (Geurts et al., 2006), LightGBM (Ke et al., 2017), and CatBoost (Dorogush et al., 2018). All of the models are tree-based ensemble approaches.

Each model is initialized with baseline configurations (e.g., random state, default parameters) and evaluated on identical training and test sets to ensure comparison regarding the results.

As **hyperparameter tuning**, we rely on the Bayesian optimization strategy (*BayesSearchCV* from `scikit-optimize`) (Frazier, 2018; Head et al., 2018). The hyperparameters `max_depth`, `n_estimators`, `learning_rate`, and `class weighting` are tuned within predefined ranges using three-fold cross-validation. To address class imbalance, *balanced accuracy* is used as the primary optimization metric, ensuring equal consideration of minority class performance. After identifying each model's hyperparameters, the final models are trained on the entire training dataset and evaluated on the test set.

Besides, we use SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) to explain and clarify the **feature importance** of the input data. SHAP measures each feature's contribution to model predictions, providing a unified and theoretically sound approach to interpreting machine learning models (Lundberg and Lee, 2017).

This analysis improves our understanding of the features that influence the classification of PV suitability.

5 RESULTS AND DISCUSSION

The results of the applied ML models are evaluated based on commonly applied classification metrics. We rely on Precision, Recall, F1 Score, Log Loss, Balanced Accuracy (BA), ROC-Auc, and Average Accuracy (AA). Given the dataset's imbalance, BA and F1 scores are prioritized to ensure robust performance across all classes. Log Loss also evaluates the reliability of probability estimates, while ROC-AUC and Average Precision provide insights into the models' discriminative abilities.

Table 2 summarizes the classification results of the ML models on the test dataset. All ML models demonstrate strong performance with high Precision, Recall, and F1 scores.

In particular, the **XGBoost** model achieves the highest balanced accuracy with 99 % on the test dataset, along with exceptional precision, Recall, and F1 score values. XGBoost exhibits the lowest log loss with 0.05, indicating excellent calibration and reliable probability estimates.

The CatBoost and Extra Trees classifiers also perform robustly, with balanced accuracy scores of 99 % and log loss values of 0.05 and 0.06, respectively. These models maintain high precision and recall, accurately classifying suitable and unsuitable parking lots.

Although the Random Forest and LightGBM models show slightly higher log loss values of 0.10 and 0.12, respectively, they still deliver commendable balanced accuracy scores of 99 %. Their strong performance underscores their effectiveness in handling class imbalance within the dataset.

Slight Recall and F1 score variations emphasize nuanced differences in each model's ability to capture relevant positive instances while maintaining overall accuracy (see Table 2).

Due to the consistent classification performance of the XGBoost classifier in terms of balanced accuracy, this model is selected for further analysis and classification of all parking lots in Hesse. Unlike other models, XGBoost efficiently handles imbalanced datasets, reducing bias toward the majority class (suitable parking lots) while maintaining high recall for the minority class (unsuitable lots). Due to its computational efficiency and handling of high-dimensional feature spaces, it is well-suited for large-scale applications, making it a practical choice for real-world urban energy planning. By confirming that XGBoost outperforms other tree-based ensemble models, we provide strong empirical justification for its selection as the optimal classifier for this study.

Figure 2 visualizes the classification results of the XGBoost model of the training and test datasets, respectively. It elucidates the number of correctly and incorrectly classified areas and the influence of various features on the model's predictions.

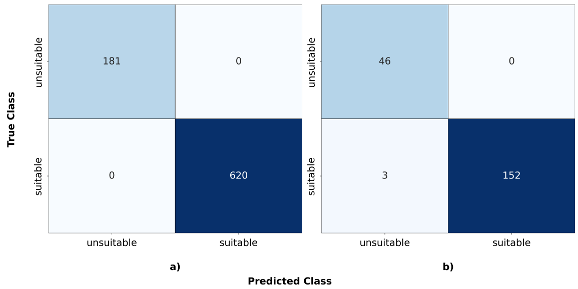


Figure 2: Confusion matrix of the XGBoost model on the (a) training dataset and (b) test dataset.

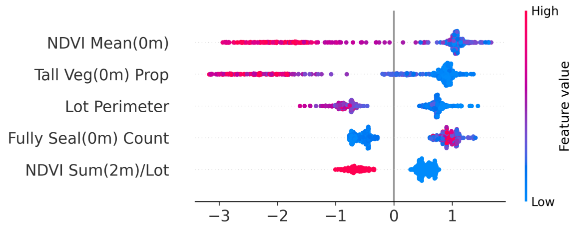


Figure 3: SHAP plot illustrating the feature importance for the XGBoost model (five most important features, listed left) based on the training dataset. Colors indicate feature values. While high values are red, low values are blue.

Figure 3 illustrates how individual input features influence the XGBoost model's classification of parking lots as suitable or unsuitable for PV installations. Each SHAP value represents the importance of a feature, with blue indicating low and red indicating high feature values. The most significant feature is the `NDVI Mean (0m)`, representing the mean vegetation index within the 0 m buffer (location of the parking lots themselves). High NDVI values (red), indicating dense vegetation, reduce the suitability for PV installations.

Another essential feature is the proportion of `tall vegetation` on the parking lot, where higher proportions (red) also decrease PV suitability. In contrast, the `fully sealed count` indicates that areas with more sealed surfaces (blue) increase suitability, as these surfaces are better suited for PV. Additional features such as the `lot perimeter` and the `NDVI sum/lot` ratio within the 2 m buffer zone show similar trends, where sealed and less vegetated areas enhance suitability.

In summary, vegetation-related features negatively impact PV suitability, while sealed surfaces contribute positively.

Table 2: Classification results of the applied ML models on the test dataset. BA means Balanced Accuracy. The best-performing model is highlighted.

ML Model	Precision	Recall	F1 Score	Log Loss	BA
Random Forest	1.00	0.96	0.98	0.10	0.98
XGBoost	**1.00**	**0.98**	**0.99**	**0.05**	**0.99**
Extra Trees	1.00	0.98	0.99	0.06	0.99
LightGBM	1.00	0.98	0.99	0.12	0.99
CatBoost	1.00	0.97	0.99	0.05	0.99

6 APPLICATION OF THE BEST MODEL

Since we aim to investigate all parking lots within Hesse, we apply the strong classification model, XGBoost, to the entire parking lot data.

6.1 Assessing the Parking Lot Dataset of the Study Region Suitable for PV

Analysis of Parking Lots Suitable for PV in Hesse: The parking lot dataset's overlap with settlement areas provides insight into the distribution of parking lots in urban and rural regions. A weighted area is calculated for each class to quantify classification uncertainty, shown as black error bars in Figure 4. This weighted area is derived by multiplying the actual area of each parking lot by the model's prediction probability for that class label. Weighted areas are summed separately for parking lots within and outside settlement areas. Their deviations are used to compute the standard deviation of the weighted areas.

The total potential amounts to 28.541 km^2 of PV suitable parking lots, with a model prediction uncertainty of 0.47 km^2. These parking lots represent approximately 0.1 % of Hesse's total area. This analysis highlights the potential for using parking lots for PV installations. However, this potential is gradually reduced when practical constraints, including technical feasibility, economic viability, legal considerations, and environmental impact, are considered.

Identifying the percentage of parking lot areas suitable for PV systems is essential to accurately and generally assess the potential of PV parking lot installations. This area coverage factor, which indicates the percentage, can vary significantly between studies, ranging from 18 % to 79.4 %. These differences are attributed to the varying assumptions and methodologies used in the existing investigation (Maier et al., 2024).

Of the 28.541 km^2 classified as suitable, 1.606 km^2 are identified as existing rooftops. Although rooftops are also suitable for PV, they are excluded from the analysis. Additionally, 1.637 km^2 of tall vegetation and 0.401 km^2 of medium-sized vegetation were subtracted, as removal of vegetation would have a significant environmental impact. Regarding these exclusions, the upper range of available parking lot area for PV is 24.921 km^2.

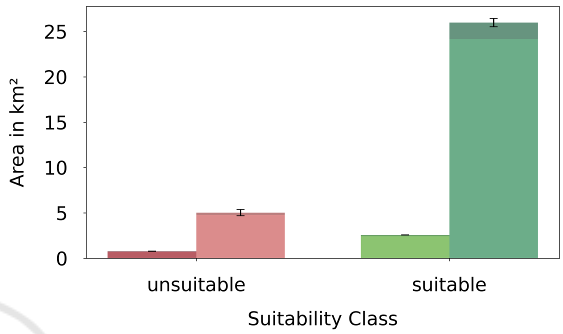


Figure 4: Distribution of parking lots in urban and rural areas on the reference dataset (dark red and green) and the entire dataset of Hesse (light red and green), classified using the XGBoost model. Shaded areas (right) represent parking lots outside urban areas. The black boxplots represent error bars, while the estimated area represents the theoretical potential of suitable PV parking lots.

To minimize environmental impact, PV installations should focus on paved surfaces, reducing the available area to 21.804 km^2 (= 50.8 % of the theoretical potential parking lot area). These figures align with previous upper estimates for parking PV potential.

By incorporating various spatial and geometric features as input for the data-driven models, we can provide explainable information on what the ML approaches rely on to solve the classification task. This *explainability* can be a first step towards more transparency, which is often lacking according to (Maier et al., 2024), concerning the overall topic of PV on parking lots.

6.2 Solar Energy Investigation of Selected Parking Lots

One key challenge in validating ML models for PV suitability assessment is the limited availability of high-resolution solar radiation data. In the case of Hesse, such data is not freely accessible for the entire region, restricting our ability to perform a comprehensive solar energy potential analysis.

To address this challenge, we have conducted a case study using 193 manually validated parking lots in Frankfurt am Main, where high-resolution solar data from the Hessian Solar Register was available. This targeted validation serves as a representative test of our model's effectiveness.

Figure 5 compares solar radiation ($\mathrm{kW\,h\,m^{-2}}$) for 193 parking lots in Frankfurt am Main, classified as suitable or unsuitable by the model. A Mann-Whitney U test confirms that parking lots predicted as suitable exhibit significantly higher solar radiation than those predicted as unsuitable ($p<0.001$, Cliff's $\Delta = -0.848$). Suitable lots show median solar radiation of $950\,\mathrm{kW\,h\,m^{-2}}$, compared to $702\,\mathrm{kW\,h\,m^{-2}}$ for unsuitable lots. These findings demonstrate that the model effectively identifies locations with higher solar potential despite having no solar radiation as an input. Further, this result shows that the manual labeling of selected parking lots as suitable and unsuitable has been conducted correctly.

While the analysis covers approximately 20 % of all parking lots in Frankfurt am Main, the results validate the suitability predictions regarding solar radiation. The classification correlates well with expected solar potential, even though our model does not explicitly use solar radiation as an input feature.

Integrating this framework with high-resolution solar radiation datasets and an economic analysis tool, such as the Hessian solar register, could optimize yield and return-on-investment analyses for PV deployment across the entire state of Hesse without requiring manual labeling.

These findings demonstrate that our feature-driven classification approach can effectively predict solar suitability without comprehensive solar radiation datasets. Future work could integrate solar modeling techniques or partner with governmental agencies to obtain broader access to radiation data for full-scale validation.

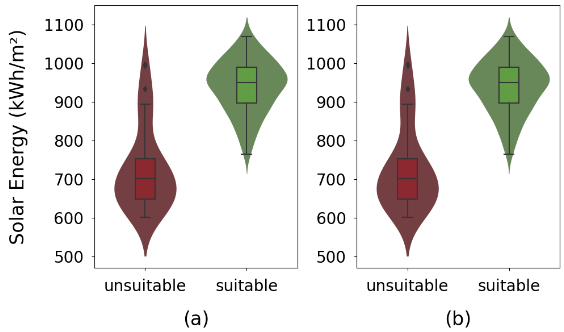


Figure 5: Violin plots comparing solar radiation for unsuitable (red) and suitable (green) parking lots classified by the XGBoost model.

7 CONCLUSIONS AND OUTLOOK

In conclusion, the innovative framework proposed establishes a solid foundation for assessing the PV potential in urban parking areas at scale, encouraging new research and applications in renewable energy and urban development. By integrating diverse datasets, such as OSM, ATKIS, and high-resolution geospatial imagery and utilizing advanced machine learning techniques, we identified and classified suitable parking areas for PV canopy installations. The model's effectiveness is illustrated through its application in Hesse, Germany. The XGBoost model achieved an impressive classification accuracy of 99 %, distinguishing suitable sites based on features like vegetation indices and sealing ratios. About $21.8\,\mathrm{km}^2$ of parking areas were identified as suitable for PV, promoting sustainable energy solutions.

Validation against solar radiation data further confirms the model's reliability without requiring explicit radiation inputs. This approach bridges the gap between theoretical potential and actionable insights, equipping urban planners and policymakers to optimize energy systems toward climate-neutrality goals. Its independence from high-resolution radiation models increases its applicability in data-scarce environments, and the automated processing enhances scalability.

Future research directions can include:

- *Assess Economic Feasibility:* Analyzing installation costs and return-on-investment to help prioritize PV deployment sites.
- *Integrate Additional Validation Datasets:* Collaborating with regional agencies for broader solar data access to evaluate classification performance.

Overall, this study offers a scalable, data-driven framework for assessing PV potential in urban settings. This has implications for urban planning, renewable energy investments, and policy-making. The method can be replicated in other regions, validating its applicability across diverse geographical and climatic contexts.

ACKNOWLEDGEMENTS

We thank the HMWVW for the part-funding of this work.

REFERENCES

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM.

Dorogush, A. V., Ershov, V., and Gulin, A. (2018). Catboost: Gradient boosting with categorical features support. *ArXiv*, abs/1810.11363.

Figueiredo, R., Nunes, P., and Brito, M. C. (2017). The feasibility of solar parking lots for electric vehicles. *Energy*, 140:1182–1197.

Frazier, P. I. (2018). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Head, T., Pak, M., Louppe, G., Shcherbatyi, I., Charras, F., Vinícius, Z., and Fabisch, A. (2018). scikit-optimize/scikit-optimize.

HMWEVW (2023). Photovoltaikpflicht für Parkplätze und Landesgebäude. Press Release.

IPCC (2022). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the IPCC*. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg3/ (accessed on 16 January 2025).

Ivanova, A., Chassin, D., Aguado, J., Crawford, C., and Djilali, N. (2020). Techno-economic feasibility of a photovoltaic-equipped plug-in electric vehicle public parking lot with coordinated charging. *IET Energy Systems Integration*, 2(3):261–272.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pages 3146–3154.

Krishnan, R., Haselhuhn, A., and Pearce, J. M. (2017). Technical solar photovoltaic potential of scaled parking lot canopies: A case study of walmart usa. *Journal on Innovation and Sustainability RISUS*, 8(2):104–125.

Landes Energie Agentur Hessen (2025). Solar Kataster Hessen: A Tool for Utilizing Solar Energy. Accessed: 2025-01-22.

Ludwig, D., Tegeler, A., Schmedes, D., Tomhave, L., Hensel, A., Forster, J., Kleinhans, A., Heinrich, M., John, R., and Schill, C. (2024). Potenziale für Photovoltaik an Bundesfernstraßen. Technical report, Bundesanstalt für Straßenwesen (BASt), Bergisch Gladbach, Germany.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4768–4777. Curran Associates Inc.

Maier, R., Lütz, L., Risch, S., Kullmann, F., Weinand, J., and Stolten, D. (2024). Potential of floating, parking, and agri photovoltaics in germany. *Renewable and Sustainable Energy Reviews*, 200:114500.

Marneni, N., Agarwal, N., and Munshi, N. (2021). Techno-economic analysis of solar carports for a commercial building: A u.s. case study. *Energy Reports*, 7:160–168. https://www.sciencedirect.com/science/article/pii/S2352484721000686 (accessed on 16 January 2025).

More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. *ArXiv*, abs/1608.06048.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). scikit-learn: Machine learning in python. https://scikit-learn.org/. Version 0.24.2.

Santamouris, M. (2020). Recent progress on urban overheating and heat island research: A review. *Current Opinion in Environmental Science & Health*, 13:89–97.

Solar Cluster BW (2022). Photovoltaik-Parkplätze: Solarüberdachung von Park- und Stellplätzen. Technical report, Solar Cluster Baden-Württemberg e.V., Stuttgart, Germany. Faktenpapier des Photovoltaik-Netzwerks Baden-Württemberg. Original authors: Solar Cluster Baden-Württemberg e.V., Gerhard Stryi-Hipp, Thomas Uhland.

Wirth, H. (2023). Aktuelle Fakten zur Photovoltaik in Deutschland. Technical report, Fraunhofer-Institut für Solare Energiesysteme ISE. Available at www.pv-fakten.de, accessed on January 20, 2025.