Masters Program in Geospatial Technologies

SUITABILITY MAPPING OF SOLAR POWER PLANTS USING AN EXPLAINABLE AI-BASED APPROACH

Mawanane Hewa Rasanka Mangala De Silva

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

SUITABILITY MAPPING OF SOLAR POWER PLANTS USING AN EXPLAINABLE AI-BASED APPROACH

Dissertation supervised by:

Leonardo Vanneschi, PhD

NOVA Information Management School

Universidade Nova de Lisboa

Lisbon, Portugal

Co-supervised by:

Marco Painho, PhD

NOVA Information Management School

Universidade Nova de Lisboa

Lisbon, Portugal

Co-supervised by:

Michael Gould, PhD

GEOTEC

Universitat Jaume I

Castellón, Spain

February 20, 2024

ACKNOWLEDGMENTS

I extend my heartfelt gratitude to my supervisor, Prof. Dr. Leonardo Vanneschi, for his steadfast support, guidance, and invaluable insights during the study. Additionally, I am deeply thankful to my co-supervisors, Prof. Dr. Marco Painho and Prof. Dr. Michael Gould, for their valuable guidance, unwavering encouragement, and consistent support throughout the study. Their expertise and motivation have played a pivotal role in shaping the trajectory of my work.

I express my genuine gratitude to the faculty and staff of Universidade Nova de Lisboa, University of Münster, and Universitat Jaume I for their provision of resources, assistance, and an environment conducive to academic exploration. Additionally, I would like to thank the Erasmus Mundus program for its funding support towards my Master of Science in Geospatial Technologies.

I am thankful to my family in Sri Lanka for their unwavering support in every aspect, propelling me forward. Last, I thank my wife, Prasadi, for her constant encouragement and steadfast support throughout the study. Your motivation has been indispensable in enabling me to complete this study successfully.

This thesis represents the culmination of the combined efforts, encouragement, and support from numerous individuals, and for this, I am sincerely grateful.

SUITABILITY MAPPING OF SOLAR POWER PLANTS USING AN EXPLAINABLE AI-BASED APPROACH

ABSTRACT

The conventional approaches to identifying optimal locations for solar power plants are traditionally handled through Multi-Criteria Decision-Making (MCDM) techniques, which often suffer from subjectivity and lack of transparency. Although recent advancements in Machine Learning (ML) have offered potential solutions to MCDM-based methods, traditional ML models struggle to explain their predictions and operate without assuming that current solar power plants are in optimal locations.

Thus, this research represents an integrated explainable AI approach with ML methods, aiming to enhance the transparency and comprehensibility of results and introduce a novel paradigm by employing the classified efficiency of existing plants, categorized into five classes, as the dependent variable for ML models, thereby challenging the prevalent assumption inherent in conventional ML models. Twelve independent variables selected through a literature review were used, along with the classified efficiency values, to train five ML models, namely, Random Forest (RF), Support Vector Machines (SVM), Multi-layer Perceptron (MLP), Decision tree (DT), and K-nearest neighbors (k-NN).

Following the assessment of the models' accuracies, the RF model, which achieved 88% overall accuracy, was subsequently explained using SHapley Additive exPlanations (SHAP), revealing Solar Radiation (13%) and Cloud Index (12%) as the most influential variables for the resulting predictions. In comparison, Aspect (5%) was identified as the least significant parameter to the model predictions.

The final solar suitability map produced with the superior RF model identified approximately 5% of the total land area in the USA as highly suitable for constructing solar power plants, ensuring optimal operational efficiency. Additionally, 55% of the land is moderately suitable for such establishments. Conversely, approximately 9.5%

of the total land area, equivalent to 766,654 km^2, is deemed permanently unsuitable for solar power plant construction.

Sustainable Development Goals (SGD):

KEYWORDS

Renewable Energy

Geographical Information Systems

Machine Learning

Explainable Artificial Intelligence

Suitability Mapping

United States of America

ACRONYMS

A	–	Aspect
AHP	–	Analytic Hierarchy Process
AI	–	Artificial Intelligence
ANN	–	Artificial Neural Network
AT	–	Air temperature
AUC	–	Area under the Curve
AUROC	–	Area Under the ROC curve
CI	–	Cloud index
CSV	–	Comma-separated values
DT	–	Decision Tree
E	–	Elevation
EIA	–	Energy Information Agency
GIS	–	Geographic Information Systems
GW	–	Gigawatt
k-NN	–	k-Nearest Neighbor
LULC	–	Land use/ land cover
MCDM	–	Multi-Criteria Decision Making
MDI	–	Mean Decrease in Impurity
ML	–	Machine learning
MLP	–	Multi-layer Perceptron
MW	–	Megawatt
PC	–	Proximity to the city center
PD	–	Population density

PG	–	Proximity to grid
PR	–	Proximity to road network
RF	–	Random Forest
ROC	–	Receiver Operating Characteristic
S	–	Slope
SHAP	–	Shapley additive explanations
SMOTE	–	Synthetic Minority Oversampling Technique
SMV	–	Support Vector Machines
SR	–	Solar radiation
Ti	–	Tolerance
TOPSIS	–	Technique for Order Performance by Similarity to Ideal Solution
USA	–	United States of America
VIF	–	Variance Inflation Factor
WS	–	Wind speed
XAI	–	Explainable Artificial Intelligence

INDEX OF THE TEXT

Page

INDEX OF TABLES

Page

INDEX OF FIGURES

Page

1. INTRODUCTION

1.1 Overview of the work

Electricity stands as one of the most crucial infrastructures in today's industrial era (Bayounis & Eldamaty, 2022). According to World Energy and Climate statistics, global energy consumption continues to grow steadily, with a 2.1% increase recorded in 2022. Despite ongoing efforts to shift away from traditional fossil fuel-based power generation methods towards renewable energy sources, according to predictions by the World Energy Council, petroleum and natural gas are projected to maintain their dominance, contributing more than half of the world's energy sources in 2040. This persistence of fossil fuels in our energy production poses a significant challenge, as it could lead to an increase in carbon emissions, intensifying global warming and deteriorating the sustainability of our planet. This underscores the urgency of accelerating the transition to renewable energy sources.

Renewable energy, as defined by (Suprova et al., 2020), encompasses energy sources derived from natural resources capable of self-regeneration within a relatively short timeframe. Prominent examples of these renewable sources include solar, wind, geothermal, and hydropower, which are widely used for energy production. Among the array of renewable energy options, solar energy stands out as the most valuable renewable energy source because of its abundant, freely accessible, and clean availability, as well as its cost-effectiveness and simplicity in energy production processes (Halder et al., 2022). Solar energy stands out as a prominent renewable energy source worldwide, with a cumulative installed power capacity of 942 GW as of 2022 (REN21, 2022). A solar power plant is a facility that captures and uses sunlight as a source of energy and makes it available as a form of electricity for human use. The efficiency of a solar power plant is a critical factor to consider, influenced by various elements such as solar radiation reaching the panels, local temperature, humidity, cloudiness, and more (Suprova et al., 2020). The solar systems' energy output depends on the variability of these parameters (Rangel-Martinez et al., 2021). Additionally, the production cost of a solar power plant is closely tied to geographic variables. Selecting a location near an existing power grid can lower infrastructure development expenses for transmitting electricity to local homes and reduce transmission losses. Favorable elevations and slopes can also cut costs associated with excavation and earthwork. Furthermore, it is essential to avoid constructing solar power plants in environmentally protected areas to mitigate concerns about land use, habitat loss, water consumption, and using hazardous materials in manufacturing (Spyridonidou & Vagiona, 2023). Hence, identifying an appropriate site for establishing a solar power plant can be complex. Engaging in thoughtful and strategic construction planning that considers geological, climatic, geomorphological, and social factors is crucial before initiating the construction process (Sun et al., 2023).

With the advent of Geographical Information Systems (GIS) software and spatial analysis techniques, GIS has become a widely adopted approach for identifying suitable locations by managing multiple spatial data sources encompassing the aforementioned influencing factors (Sun et al., 2023). GIS tools possess the capability to integrate diverse spatial data sources, including maps, aerial photographs, and satellite images, as well as quantitative, qualitative, and descriptive information databases. This integrated approach significantly simplifies the storage, analysis, and visualization of spatial relationships and connections, facilitating effective solutions for decision-making challenges. The Multi Criteria Decision-making Method (MCDM) stands out as one of the most frequently utilized decision-making approaches within GIS for assessing locational suitability. Often, this method is combined with various weighting techniques such as Analytical Hierarchy Process (AHP) (Halder et al., 2022), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Wang et al., 2022) and Fuzzy Logic (Yousefi et al., 2018) in previous research studies. However, the selection of parameters to represent influencing factors and the scoring of these factors can be influenced by the researcher's subjectivity, expertise level, and practical experience in the field (Sachit et al., 2022). As a result, the evaluation outcomes often suffer from subjectivity, leading to biases in real-world scenarios. This challenge complicates the task of comparing the results generated by these methods.

Subsequently, the advent of Machine Learning (ML) methods that involve training algorithms with sample data emerged as an alternative approach for suitability modeling, allowing for the mapping of spatial suitability and hazard susceptibility (Xu et al., 2022). ML models have found successful applications in suitability mapping for various purposes, including areas such as wind and solar power plant siting (Sachit et al., 2022), hospital locations (Almansi et al., 2021), dam construction (Al-Ruzouq et al., 2019), and the development of solid waste management systems. However, as noted by (Sachit et al., 2022), traditional ML models often struggle to provide easily understandable explanations for their predictions or decisions. This highlights the importance of Explainable Artificial Intelligence (XAI), which aims to enhance transparency and interpretability in ML models, enabling a more comprehensive understanding of the impact and significance of each input criterion on the outputs of the ML model.

1.2 Research gap

Despite the significance of XAI in the field of ML, there has been a scarcity of prior research that integrates XAI with ML models specifically designed for assessing the suitability of locations for solar power plants. To the best of our knowledge, only two studies conducted by (Sachit et al., 2022) and (Sun et al., 2023) have been dedicated to integrating the XAI approach into ML for suitability mapping in the context of solar farms. However, both of these studies have assumed that the existing PV power station used as input to the ML model is in an ideal geographical setting. Interestingly, the existing literature lacks any exploration into the possibility of using electricity generation from power plants as a viable metric for assessing their efficiency.

1.3 Objectives

Consequently, to address the above-mentioned knowledge gaps, this study aims to develop an Explainable AI-based Machine Learning model for mapping the suitability of solar power plant locations with a particular focus on efficiency, which is calculated by the individual generation and the capacity of the power plants. The secondary objectives were designed to address the following questions:

1. What are the key factors that significantly influence the suitability of solar power plant site selection in different regions in the USA?
2. Which ML-based classification method is the most accurate for creating a suitability map to identify potential solar power sites using existing solar productivity data?
3. What is the influence of each factor on the final suitability map?

The results of this research have the potential to provide valuable insights for decision-makers involved in urban planning, energy policy development, and environmental conservation initiatives. Doing so can promote a cleaner, more sustainable environment and a greener future for future generations.

1.4 Thesis organization

This thesis comprises a total of six subsequent chapters. Chapter 1 is an introductory chapter, providing an overview of the research, emphasizing the research gap, and outlining the project's objectives. Moving on to Chapter 2, a comprehensive review of the existing literature is conducted, focusing on techniques related to suitability mapping for solar power plants and ML and explainable AI-based methods. Chapter 3 delves into the study area, explaining the data sets employed and detailing the methodology and tools utilized in the research. Chapter 4 is dedicated to presenting the research findings, organized with a comprehensive analysis and discussion of the obtained results. Finally, Chapter 5 serves as a concluding chapter, consolidating the research by summarizing the key findings, addressing research questions, and offering recommendations for future endeavors.

2. LITERATURE REVIEW

In the mission of reducing carbon emissions and conserving finite energy resources, such as fossil fuels, which are consumed at a rate higher than their natural replenishment, numerous countries around the world have strategized the adoption of alternative renewable energy sources as a means of meeting the energy demands of the 21st century (Hernandez, 2018). Among these nations, those located in arid and semi-arid climatic regions have exhibited a particular interest in consuming sunlight as a resource due to its abundant and freely available nature (Halder et al., 2022). Given the complicated relationship between the location of solar power plants, their efficiency, and the associated construction costs, the practice of suitability mapping for such installations has gathered substantial attention (Spyridonidou & Vagiona, 2023). This literature review synthesizes existing research on the subject, specifically focusing on using explainable AI techniques to determine the most suitable locations for solar power plants.

2.1 Solar suitability mapping

Solar suitability mapping involves the identification of geographically optimal regions for establishing solar energy generation infrastructure (Suprova et al., 2020). It comprehensively analyzes various influential factors, such as solar irradiance levels, land availability, climatic conditions, and economic viability (Maher et al., 2015a; Rida et al., 2017; Teruel-Solano et al., 2013). Changes in these factors can result in some regions on Earth's surface being more or less suitable for establishing solar power plants (Alami Merrouni et al., 2018). Typically, the existing literature reveals that areas with high solar irradiance levels, minimal environmental limitations, and proximity to established power infrastructure are favored for solar power plant installations (Asadi et al., 2023; Asadi & Pourhossein, 2021; Bayounis & Eldamaty, 2022; Halder et al., 2022). Nonetheless, the precise criteria for suitability can differ across studies, mirroring the variations in local conditions and priorities (Suprova et al., 2020). Identifying these suitable regions is a complex undertaking that requires the simultaneous evaluation and comparison of multiple influencing factors through a systematic procedure (Halder et al., 2022; Hernandez, 2018; Husein et al., 2023).

Following the introduction of Geographical Information Systems (GIS), traditional research methodologies have utilized GIS-based Multi Criteria Decision Making (MCDM) technologies to address these inquiries (Suprova et al., 2020). This process entails the creation of distinct spatial layers that represent various influencing factors, subsequently converted into spatial decisions. A weighted overlay is then applied to these layers by assigning specific weights to each input. The determination of these weights involves evaluating the significance of the factors using a variety of techniques, such as Analytical Hierarchical Process (AHP) (Halder et al., 2021), (Alami Merrouni et al., 2018), (Maher et al., 2015b), Technique for Order of Preference by Similarity to ideal Solution (TOPSIS) (Sarkodie et al., 2022), and Boolean fuzzy logic models (Yousefi et al., 2018), as well as the expert opinions from the field. Table 2.1 summarizes the methodologies used in research that apply MCDM technologies to map the suitability of solar power plants.

	Methods applied	**References**
1	Data Envelopment Analysis (DEA) and Grey Based Multiple Criteria Decision Making (G-MCDM)	(Wang et al., 2022)
2	Linear Regression Modelling (LRM) and GIS + AHP	(Asadi et al., 2023)
3	AHP + Sensitivity analysis (SA)	(Rida Azmi, Hicham Amar, 2017)
4	GIS + MCDM	(Islam et al., 2022), (Gherboudj & Ghedira, 2016), (Villacreses et al., 2022)
5	Technique for order of preference by similarity to ideal solution (TOPSIS), Complex proportional assessment (COPRAS), Multi-objective optimization on ratio analysis (MOORA) + Spearman correlation coefficient (SCC)	(Sarkodie et al., 2022)
6	MCDM + TOPSIS	(Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013)
7	MCDM	(Nzelibe et al., 2022), (Akkas et al., 2017)
8	MCDM +AHP	(Munkhbat & Choi, 2021), (Tisza, 2014), (Watson & Hudson, 2015), (Bayounis &

		Eldamaty, 2022), (Watson & Hudson, 2015)
9	GIS + AHP	(Halder et al., 2021), (Alami Merrouni et al., 2018), (Maher et al., 2015b)
10	GIS + MCDM +AHP	(Halder et al., 2022)
11	GIS + Boolean and fuzzy model	(Yousefi et al., 2018)

Table 2.1: MCDM based Solar power plant suitability methods used in previous studies

Nonetheless, the results of these researches depend upon various factors, such as the selection of criteria, the assignment of relative weights to the selected criteria, and the particular pairs of criteria used for assessing their relative importance (Asadi & Pourhossein, 2021; Husein et al., 2023; Sánchez-Lozano et al., 2015). Therefore, the conclusions drawn from these weight-based techniques are subjective and susceptible to the researcher's personal bias, expertise, and practical experience in the field (Sun et al., 2023a).

To illustrate this variability, consider two studies conducted to assess the suitability of solar power plant locations in Morocco, namely, (Taoufik et al., 2021) and (Alami Merrouni et al., 2018). These studies identified 11 and 8 influencing factors, respectively. Due to differing judgments regarding the selection of influencing factors, the two studies assigned varying degrees of influence to their primary factor, global horizontal solar radiation. In Taoufik's methodology, this factor held a 26% influence on solar site selection when applying the AHP method. In contrast, Alami concluded that direct normal irradiation was more than 50% significant in the AHP model. In addition, these two studies used different levels of importance for the influencing factors. For example, Taoufik considered solar radiation above 5 kW/m² highly suitable, while Alami considered 2.1 kW/m² the highest suitable range. Therefore, the findings of these two studies deviate significantly and cannot be readily compared to determine the most accurate results for decision-making processes. As such, the choice of parameters and scoring methods greatly influences project outcomes and is subject to the researcher's biases, level of knowledge, and real-world experience in the field. Due to these limitations and advancements in Artificial Neural Networks (ANN), researchers have adopted ML models to tackle location-based challenges (Ghimire et al., 2022; Rangel-Martinez et al., 2021).

2.2 Artificial neural networks-based approaches for solar suitability mapping

The popularity of ANN, inspired by human biological neurons, has increased as a tool for data analysis in various applications (Singh, 2019), including suitability mapping in GIS. Machine learning is a subfield of artificial intelligence that involves the development of mathematical or statistical algorithms that can learn from and make predictions or decisions based on existing data about a specific real-world inquiry (Saranya & Subhashini, 2023). The utilization of machine learning models to assess locational suitability has demonstrated a successful history of methodologies across various fields in recent years, including predicting suitable sites for dams, hospitals, solid waste landfills, and renewable energy sites. These studies have used several machine learning techniques, summarized in Table 2.2, including decision trees, Random Forests, support vector machines, and neural networks.

	Methods applied	**References**
1	GIS + Machine learning (Fuzzy membership + Fuzzy logic model)	(Haifaa Nasser Hussein, Noor Hashim Hamed, 2023)
2	ML (Hybrid deep CNN-SVR algorithm)	(Ghimire et al., 2022)
3	GIS + ML(ANN)	(Oyewola et al., 2022)
4	MCDM + ANN	(Asadi & Pourhossein, 2021)
5	ML (ANN, Support vector regression (SVR), and Gaussian Process Regression (GPR))	(Sharifzadeh et al., 2019)
6	ML (Gradient Boost Decision Tree (GBDT) and eXtreme Gradient Boosting (XGBoost)	(Hou et al., 2023)
7	ML (RF, DT, SVM, k-NN, ANN)	(Shahab & Singh, 2019)
8	GIS + ML + XAI	(Sachit et al., 2022a)
9	ML (RF, MLP, Xtreme Gradient Boosting (XGBoost) + XAI	(Sun et al., 2023b)

Table 2.2: ML based solar power plant suitability methods used in previous studies

For instance, prior research conducted by Shahab examines the effectiveness of five machine learning algorithms (Random Forest(RF), Decision Tree (DT), Support Vector Machines (SVM), K-Nearest neighbors (k-NN), and Artificial Neural Network (ANN)) for classifying the suitability of renewable energy sites based on geophysical data sets (Shahab & Singh, 2019). The study's findings revealed that the RF (F1 score of 92%) algorithm demonstrated the most remarkable performance, whereas the SVM (F1 score of 85%) showed the least accurate results. Another study conducted by Asadi employed a combination of MCDM and Multi-Layer Perceptron (MLP) to determine the suitability of solar power plant locations in Iraq (Asadi & Pourhossein, 2021). The study's findings further validate the suitability of machine learning-based methods, offering accurate and stable results for implementing global scoring capabilities in suitability analysis tasks. However, the assessment of the superiority of machine learning-based methods over traditional MCDM-based approaches needs to be addressed.

2.3 Machine learning over MCDM-based methods

A study conducted by Saha examined the effectiveness of MCDM-based AHP and Fuzzy Complex Proportional Assessment (FCOP-RAS) methods over two machine learning approaches: The Random Forest (RF) algorithm and the Multilayer Perceptron (MLP) for the suitability analysis task using fifteen influencing criteria collected through ground surveys. The results of the study showed that the accuracy of the models developed using machine learning based model achieved higher AUC scores compared to the other methods, with AUC scores of 0.947 for RF, 0.923 for AHP, 0.928 for FCOP-RAS, and 0.932 for MLP. Additionally, the study highlighted the time-saving attribute of the machine learning-based method over MCDM-based approaches due to the lower human interaction for task execution (Saha & Mondal, 2022).

Additionally, a recent study conducted a state-of-the-art analysis of ML-based and MCDM-based methods for decision support systems revealed that the ML-based decision-making systems provide robust and efficient solutions to complex problems by effectively handling large and complex datasets compared to the traditional MCDM-based methods. Moreover, the study also highlighted the ability of ML algorithms to learn from data patterns and make predictions without explicit programming, thus reducing the subjective judgments on those decision-making studies (Ali et al., 2023).

Furthermore, a study conducted by Asadi has highlighted one of the significant disadvantages of MCDM-based approaches known as local scoring property, which refers to the limitation, wherein scores in MCDM methods are derived from a restricted set of local observations, resulting in a lack of robustness and susceptibility to changes in scores when new candidates are introduced. As a solution to this issue, Asadi and colleagues introduced a novel ML-based approach using Multilayer Perceptron (MLP), which offers robust and global scoring solutions for modeling wind/solar farm siting in East Azerbaijan. This study further confirms the advantages of ML-based methods over the traditional MCDM-based methods in the context of suitability analysis projects. However, ML-based methods are proven to be more accurate and efficient, and the availability of accurately labeled ground truth samples is an essential factor for the successful training of machine learning models (Ghimire et al., 2022), (Singh, 2019), (Sachit et al., 2022a).

Additionally, the studies that used the ML-based method for locational suitability mapping have achieved accurate results and are capable of solving the issues inherited with the traditional MCDM-based methods; they often fail to explain the reason behind the output results or location selections (Sachit et al., 2022a). These studies usually rely on accuracy measures such as area under the curve (AUC) or accuracy under the receiver operating characteristics (AUROC) calculated for the validation data as the primary basis for the selections, without providing insights, such as whether the chosen regions have higher insolation or lower slopes, or vice versa (Pranav R, Shashank T K, n.d.; Saranya & Subhashini, 2023). As a potential solution, Explainable AI-based approaches have been incorporated into location suitability studies to provide a transparent understanding of how each input factor influences and contributes to the AI model's outputs.

2.4 Explainable AI in solar suitability mapping

Explainable Artificial Intelligence (XAI) is a concept pioneered by the Defense Advanced Research Projects Agency (DARPA) to explain the internal process of an AI model (Saranya & Subhashini, 2023). It involves providing user-understandable explanations for the method, the process, and the output of an ML model. It simply converts the Blackbox nature of an ML model to a Whitebox (Gohel et al., 2021). The explanations provided by XAI are categorized into two components: Knowledge-driven XAI and Data-driven XAI. Knowledge-driven XAI is employed to acquire insights into the methodology and techniques used. At the same time, data-driven XAI helps understand the impact and contribution of each input feature on the AI models' outcomes. A study conducted by Saranya has identified a wide range of application domains where XAI has been employed in past research, including healthcare, agricultural planning, finance, forecasting, social media, and more (Saranya & Subhashini, 2023). Sachit acknowledged that their study marked the initial attempt to employ an Explainable AI-based approach in mapping spatial suitability for global wind and solar systems to evaluate the specific contribution of each input criterion within the ML model (Sachit et al., 2022a). The study highlights compelling scientific evidence supporting machine learning models' robustness, accuracy, and applicability in suitability analysis projects. The performance metrics of the chosen models, including Random Forest(RF), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM), demonstrated an accuracy exceeding 70% across all models. Notably, RF emerged as the most sensitive and specific algorithm, achieving a sensitivity of 0.88 and a specificity of 0.91. The algorithm's overall accuracy and kappa coefficient were 0.90 and 0.78, respectively, while the Area Under the Curve reached 0.96. These results underscore the effectiveness of ML models, mainly RF, in enhancing the precision and reliability of suitability analyses. Further, the study also utilized the global and local explanations of the SHapley Additive exPlanations (SHAP) method to explain the ML model. The study's findings revealed that factors such as distance from city centers and temperature significantly influenced suitability decisions within the ML model.

A separate study conducted by Sun modeled the location choice of large-scale solar photovoltaic power plants in China using interpretable machine-learning techniques. This research revealed that the two most critical predictors of suitability for solar photovoltaic installation locations were consistently the vegetation index and the distance to the power grid among the selected 21 geospatial conditioning factors (Sun et al., 2023a). Nevertheless, it is essential to note that both of these studies operated assuming that the current solar power plants are situated in the most suitable areas within their respective regions (Sachit et al., 2022a; Sun et al., 2023a). However, this assumption may not hold in many instances. The literature mentioned above has also shown that using XAI techniques in solar power plant selection projects is still relatively under-researched. Only a few studies have been carried out in the past that integrated the spatial locations of the existing plants for the ML model (Sun et al., 2023a). To the best of our knowledge, no studies considered the efficiency of previous facilities for the ML model.

2.5 Criteria and restriction factors

The careful selection of criteria and restriction factors is a crucial aspect of a solar power plant suitability mapping project, as it directly influences the accuracy and relevance of the results (Suprova et al., 2020). The efficiency of solar power plants is influenced by the environmental and module features of the location and the equipment deployed in the solar power plants (Sahin et al., 2023). Table 2.3 and Table 2.4 summarizes the criteria and restriction factors in the previous studies examined in this research.

	Criteria/ Influential factor	References
1	Solar radiation	(Wang et al., 2022), (Asadi et al., 2023), (Bayounis & Eldamaty, 2022), (Gherboudj & Ghedira, 2016), (Shahab & Singh, 2019), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Nzelibe et al., 2022), (Munkhbat & Choi, 2021), (Tisza, 2014), (Halder et al., 2022), (Alami Merrouni et al., 2018), (Khandakar et al., 2019), (Akkas et al., 2017), (Sachit et al., 2022a), (Rida Azmi, Hicham Amar, 2017), (Sahin et al., 2023), (Watson & Hudson, 2015), (Sun et al., 2023b), (Yousefi et al., 2018), (Hernandez, 2018)
2	Air temperature	(Gherboudj & Ghedira, 2016), (Shahab & Singh, 2019), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Munkhbat & Choi, 2021), (Halder et al., 2021), (Khandakar et al., 2019), (Oyewola et al., 2022), (Sachit et al., 2022a), (Rida Azmi, Hicham Amar, 2017), (Sahin et al., 2023), (Hernandez, 2018)
3	Wind speed	(Wang et al., 2022), (Asadi et al., 2023), (Gherboudj & Ghedira, 2016), (Villacreses et al., 2022), (Khandakar et al., 2019), (Asadi & Pourhossein, 2021), (Sachit et al., 2022a), (Sahin et al., 2023), (Watson & Hudson, 2015)
4	Precipitation	(Wang et al., 2022), (Oyewola et al., 2022)
5	Relative humidity	(Wang et al., 2022), (Gherboudj & Ghedira, 2016), (Khandakar et al., 2019), (Oyewola et al., 2022), (Sahin et al., 2023), (Yousefi et al., 2018)
6	Sunshine hour	(Wang et al., 2022), (Akkas et al., 2017), (Sun et al., 2023b)
7	Slope	(Asadi et al., 2023), (Bayounis & Eldamaty, 2022), (Islam et al., 2022), (Shahab & Singh, 2019), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Nzelibe et al., 2022), (Munkhbat & Choi, 2021), (Halder et al., 2021), (Alami Merrouni et al., 2018), (Asadi & Pourhossein, 2021),

		(Sachit et al., 2022a), (Rida et al., 2017), (Sun et al., 2023b), (Yousefi et al., 2018), (Hernandez, 2018)
8	Cloud Index	(Wang et al., 2022), (Sachit et al., 2022a)
9	Technical information and assistance	(Wang et al., 2022)
10	Geology	(Wang et al., 2022)
11	Skilled manpower availability	(Wang et al., 2022)
12	Electricity demand	(Wang et al., 2022), (Sun et al., 2023b)
13	Land Costs	(Wang et al., 2022), (Sun et al., 2023b), (Sun et al., 2023b)
14	Local resident attitude	(Wang et al., 2022)
15	Government policies and laws	(Wang et al., 2022), (Sun et al., 2023b)
16	Land use/ land cover	(Wang et al., 2022), (Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Nzelibe et al., 2022), (Tisza, 2014), (Halder et al., 2021), (Oyewola et al., 2022), (Asadi & Pourhossein, 2021), (Sachit et al., 2022a), (Rida Azmi, Hicham Amar, 2017), (Sun et al., 2023b), (Yousefi et al., 2018)
17	Support mechanisms	(Wang et al., 2022)
18	Wildlife and endangered species impact	(Wang et al., 2022)
19	Harmful toxin emission	(Wang et al., 2022)
20	Energy saving benefits	(Wang et al., 2022)
21	Transmission grid accessibility	(Wang et al., 2022), (Asadi et al., 2023), (Bayounis & Eldamaty, 2022), (Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Nzelibe et al., 2022), (Munkhbat & Choi, 2021), (Tisza, 2014), (Alami Merrouni et al., 2018), (Sachit et al., 2022a), (Rida Azmi, Hicham Amar, 2017), (Watson & Hudson, 2015), (Sun et al., 2023b), (Hernandez, 2018)
22	Proximity to road network	(Wang et al., 2022), (Asadi et al., 2023), (Bayounis & Eldamaty, 2022), (Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Munkhbat & Choi, 2021), (Tisza, 2014), (Halder et al., 2021), (Alami Merrouni et al., 2018), (Asadi & Pourhossein, 2021), (Sachit et al., 2022a), (Rida Azmi, Hicham Amar, 2017), (Watson & Hudson, 2015), (Sun et al., 2023b), (Yousefi et al., 2018), (Hernandez, 2018)
23	Residential areas/ Population density	(Wang et al., 2022), (Nzelibe et al., 2022), (Alami Merrouni et al., 2018), (Sachit et al., 2022a), (Watson & Hudson, 2015), (Sun et al., 2023b)
24	Proximity to city/ urban	(Asadi et al., 2023), (Bayounis & Eldamaty, 2022), (Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Villacreses et al., 2022), (Asadi &

		Pourhossein, 2021), (Sachit et al., 2022a), (Rida et al., 2017), (Yousefi et al., 2018)
25	Aspect	(Bayounis & Eldamaty, 2022), (Munkhbat & Choi, 2021), (Halder et al., 2021), (Rida Azmi, Hicham Amar, 2017), (Sun et al., 2023b)
26	NDVI	(Gherboudj & Ghedira, 2016), (Shahab & Singh, 2019), (Sun et al., 2023b)
27	Atmospheric water vapor content	(Gherboudj & Ghedira, 2016)
28	Elevation	(Gherboudj & Ghedira, 2016), (Shahab & Singh, 2019), (Villacreses et al., 2022), (Nzelibe et al., 2022), (Munkhbat & Choi, 2021), (Tisza, 2014), (Halder et al., 2021), (Oyewola et al., 2022), (Sachit et al., 2022a), (Sun et al., 2023b), (Yousefi et al., 2018)
29	Dust/ aerosol optical thickness/ PM10 content	(Gherboudj & Ghedira, 2016), (Khandakar et al., 2019), (Sun et al., 2023b)
30	Soil type	(Halder et al., 2021)
31	PV surface temperature	(Khandakar et al., 2019), (Sahin et al., 2023)
32	Air pressure	(Asadi & Pourhossein, 2021), (Sahin et al., 2023)
33	Natural disasters	(Sachit et al., 2022a)
34	Module power	(Halder et al., 2021), (Sahin et al., 2023)
35	Gross GDP	(Sun et al., 2023b)

Table 2.3: Criteria used in Solar power plant suitability studies

Restriction factor	**References**
Protected areas/ Areas of high landscape value/ Archaeological sites / Paleontological sites/ Cultural heritage/ Areas of special protection for birds (ZEPAs)/ Community Interest Sites (LICs)/ Cattle trails / Wildfire destinations	(Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Tisza, 2014), (Halder et al., 2021), (Asadi & Pourhossein, 2021), (Watson & Hudson, 2015), (Hernandez, 2018)
Water bodies/ Watercourses and streams/ Water infrastructure	(Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Tisza, 2014), (Halder et al., 2021), (Alami Merrouni et al., 2018), (Rida Azmi, Hicham Amar, 2017), (Sun et al., 2023b), (Yousefi et al., 2018)
Cities/ Urban Lands	(Islam et al., 2022), (Sánchez-Lozano et al., 2015), (Teruel-Solano et al., 2013), (Villacreses et al., 2022), (Tisza, 2014)
Transportation networks	(Sánchez-Lozano et al., 2015)
Military zones	(Sánchez-Lozano et al., 2015)
Mediterranean coast	(Sánchez-Lozano et al., 2015)
Mountains	(Sánchez-Lozano et al., 2015)

Table 2.4: Restriction factors used in Solar power plant suitability studies

2.6 Renewable energy in the United States of America

The United States of America (USA) is among the largest economies in the world, with a substantial demand for electricity to meet the needs of its large population and industrial/commercial consumption (Tisza, 2014). As per the Energy Information Agency (EIA), in 2022, the USA consumed 100.41 quadrillion BTUs (British Thermal Units) of electricity, representing a 3% increase from 2021, with further increases expected in subsequent years (EIA, 2023). Following President Biden's executive order to "Tackle the Climate Crisis at Home and Abroad," the USA plans to install 30 GW of solar power in 2030. With this forecast, USA solar capacity has grown by an average of 24% annually over the last decade from 13,751 MW installed capacity in 2013 to 145,115 MW in 2022 (U.S. Renewable Energy Factsheet, 2023). Hence, creating novel techniques for identifying appropriate and efficient land parcels for renewable energy sites appears to be an essential and timely initiative for the USA.

Hence, in light of the identified research gap and the suitability of the USA as a potential region for analysis, the current study aims to employ the categorized efficiency of existing solar power plants, determined based on their electricity generation and capacity, as the dependent variable within the machine learning model. The study further considered 12 geospatial factors (including Solar radiation, Air temperature, Cloud index, Wind speed, Elevation, Slope, Aspect, Land cover land use, Proximity to the city center, Proximity to the electricity grid, Proximity to roads, and Population density) specifically selected for the study area as independent variables, prior to integrating these independent factors into the ML model. Before integrating those independent factors into the machine learning model, the factors underwent a correlation assessment to minimize data redundancy, thus enhancing the model's predictive accuracy.

3. DATA AND METHODOLOGY

This chapter delves into the topics of the study area, explaining the data sets employed and detailing the methodology and tools utilized in the research.

3.1 Study area and Data

The current study was conducted within the continental United States of America, as illustrated in Figure 3.1, excluding the states of Alaska and Hawaii. The study used various open-source datasets, as outlined in Table 3.1. After acquiring the necessary initial data, several pre-processing steps were executed to generate the final geodatabase, as described below.

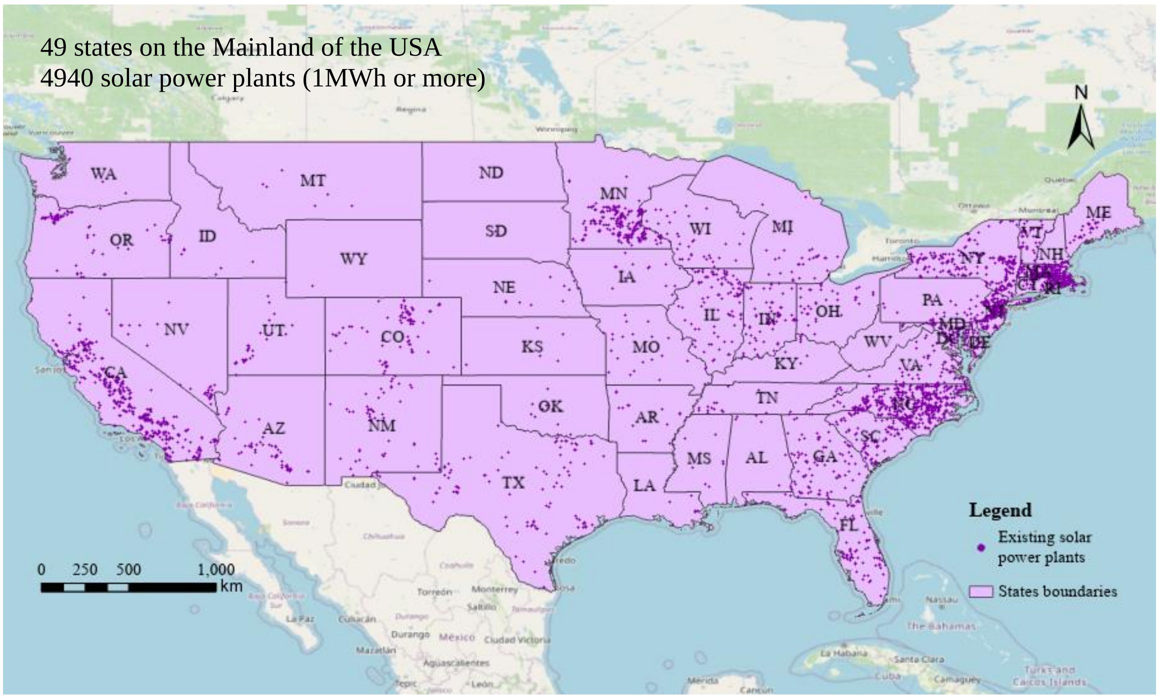

Figure 3.1: Study area

Data set	Description	Year of data set used	Source
Solar power plant location	Geographical locations of the existing solar power plant locations	2022	Open infrastructure Maps
Solar power generation and capacity	Annual solar power generation and capacity of solar plants	2022	Energy Information Administration
Administrative boundary	Administrative boundaries of the united states of America	2015	Diva GIS
Solar radiation	Accumulated global horizontal irradiation	2022	Solar Atlas 2.9
Air temperature	Mean annual air temperature	2022	Solar Atlas 2.9
Cloud Index	Mean annual cloud cover	2022	Earth Env
Wind speed	Mean annual wind speed	2022	Global wind atlas 3.3
Elevation	Digital elevation model	2022	Earth Env
Slope	Slope	2022	Earth Env
Land use land cover	Land use of the study area is divided into 20 land use classes	2021	USGS
World cities and towns	Geographical locations of the main cities in the USA	2023	Simple Maps
Electric Power Transmission Lines	Electric transmission lines in the USA	2023	US energy atlas
Road network	Road network	2020	Data.GOV
Population density	Population density map	2020	Humanitarian Data
Protected areas	Environmentally reserved areas include reservoirs, forests, national parks, marine protected areas, etc.	2023	Protected planet

Table 3.1: Necessary initial data

3.2 Methodology

The methodology proposed in this study is depicted in Figure 3.2, which comprises four distinct sections: data pre-processing, ML model selection, XAI for model description, and the creation of the final suitability map for the target area. Each section is explained in the subsequent parts of the chapter to provide a detailed understanding of the process.

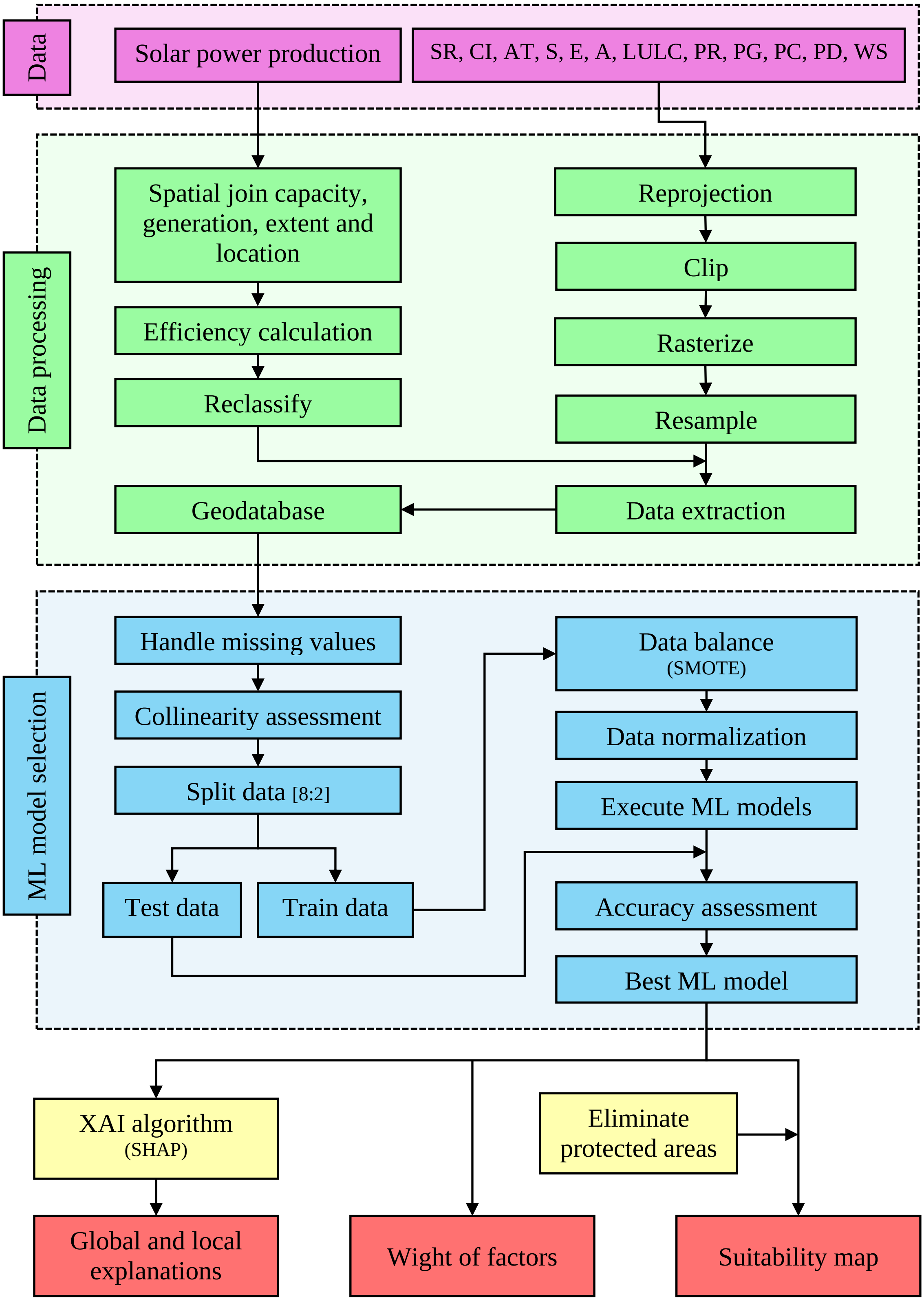

Figure 3.2: Proposed methodology

3.2.1 Data pre-processing

Data pre-processing is a vital step in ML projects. The primary purpose of data pre-processing is to transform or encode the data set used in the application so that the characteristics of the data can be identified and easily interpreted by the learning algorithm (Sirefelt Rickard, 2004). Hence, the first procedural step encompassed converting the dataset into a geodatabase optimized for the ML models.

To initiate data pre-processing, the USA Contiguous Albers Equal Area Conic coordinate system was employed to project all geographical layers clipped to the study area layer. After calculating the aspect of the study region using ArcGIS-built tools for the elevation data set, raster layers for each variable were calculated utilizing the vector-to-raster conversion tool, and all layers were resampled to a 1km spatial resolution. Given that solar power generation and plant capacity data were obtained in CSV format, the geographic locations of the solar power plants were used to join these attributes to the solar power locations spatially.

The next step involved obtaining a location within each solar power plant polygon with a capacity higher than 1 megawatt-hour and extracting cell values at those locations of solar power facilities to generate the final geodatabase. This database comprised 12 independent variables, including Solar Radiation (SR), Air Temperature (AT), Aspect (A), Proximity to City centre (PC), Cloud Index (CI), Elevation (E), Proximity to Grid line (PG), Land Use Land Cover (LULC), Population Density (PD), Proximity to the Road network (PR), Slope (S), and Wind Speed (WS) and a dependent variable representing the efficiency of the power plants.

In the subsequent phase of the process, a new variable representing the efficiency class of solar power plants was determined using the described mechanism below. The module power or the capacity of a solar power plant signifies the highest amount of solar energy supplied by the photovoltaic system under standard test conditions. It is directly linked to the energy generation of the resource, with the solar power plant operating at peak efficiency when it produces energy equal to its capacity (Sahin et al., 2023). As the dataset included the hourly capacity of the solar power plants and their annual generation, the efficiency of the solar power plants was calculated using Formula 3.1.

$$Effieciency = \frac{Annual\ net\ generation}{365\ days * 24\ hours * capacity} * 100\% \quad (3.1)$$

Due to the lack of existing literature providing cut-off values for efficiency categories in solar power plants, hypothetical assumptions were made for classification as outlined below.

If the efficiency exceeded 50%, it indicated that the power plant operated optimally for at least 12 hours a day. Based on this, power plants with an efficiency value of more than 40% were categorized as those located in the most suitable geographical areas, and the power plants with efficiencies below 1% were considered indicatives of the permanently unsuitable regions. Subsequently, the remaining solar power plants were categorized into three classes based on their efficiency levels. Power plants with efficiencies ranging from 1 to 14 were classified as marginally not suitable, those with efficiencies from 14 to 27 were classified as marginally suitable, and those with efficiencies from 27 to 40 were classified as moderately suitable. The upper value of each class was not included in the range. Once the geodatabase was established, the subsequent processing steps were implemented to render it suitable for the ML model.

- **Handling missing values:**

Missing values in the geodatabase were addressed by replacing them with the mean and modes of the available data, utilizing the mode for land use, land cover, aspect, population density, and the mean for the remaining variables.

- **Collinearity and Multicollinearity assessment:**

In order to mitigate the influence of redundant variables on the ML model, an evaluation of collinearity and multicollinearity was undertaken. This involved generating a correlation matrix and computing the variables' Tolerance (Ti) and Variance Inflation Factor (VIF). The analysis signals multicollinearity concerns when T is below 0.1, and VIF exceeds 10. Additionally, collinearity issues are identified when the correlation approaches ± 1.

- **Data splitting:**

The dataset was then split into two classes for the training and testing datasets, following the ratio of 8:2, consistent with other studies (Sachit et al., 2022a), (Sun et al., 2023a). The ML models were trained using the training dataset, and the accuracy of the models was assessed using the testing dataset.

- **Data balancing:**

Due to the unequal number of observations across efficiency classes, the Synthetic Minority Oversampling Technique (SMOTE) was employed to duplicate observations in the minority class, ensuring the impartiality of the ML models.

- **Data normalization (Training data set) and Rescaling (Testing data set):**

Given the wide variation in units among the independent variables, a minimum-maximum normalization method was utilized to standardize the training data variation from 0 to 1 for each variable using the calculated minimum and maximum values obtained exclusively for the training data set. Then, the calculated minimum and maximum values were used to rescale the testing data set to reduce the aforementioned wide variations of the testing data set.

3.2.2 Machine learning models

ML models are designed to learn patterns and relationships from labeled data, enabling them to make predictions on new, unseen data. The two main types of ML models are supervised learning and unsupervised learning. Supervised learning models, including Random Forest, Support Vector Machines, and Multi-layer Perceptron, are particularly effective in addressing classification and regression challenges. On the other hand, unsupervised learning methods are employed for clustering tasks. The present study utilized five supervised ML models: Random Forest Classifier, Decision Tree, Multi-layer Perceptron, Support Vector Machines, and K-Nearest Neighbors to categorize the suitability of solar power plants into five classes as permanently unsuitable, marginally not suitable, marginally suitable, moderately suitable, and highly suitable. The classification methods have demonstrated higher accuracies for classification tasks in previous studies. The Models were developed in a Python environment using the Scikit-learn package.

3.2.2.1 Random Forest classifier (RF)

A Random Forest classifier is an enhanced version of a decision tree where numerous uncorrelated decision trees are combined into a single structure called a forest. It is widely used to address the classification and regression application due to its ability to classify the input samples of high dimensional variables without dimension reduction or overfitting (Sun et al., 2023b). The selection of the number of trees for this classifier is a complex task that needs to be carefully conducted to prevent unnecessary noises and overfitting of the model. The Random Forest model used in this study was developed with 1000 uncorrelated trees to obtain the most accurate results for the classification problem.

3.2.2.2 Support Vector Machines (SVM)

SVM is a robust classifier built on the theoretical concepts of statistical learning to find the optimal hyperplane in a high-dimensional feature space to separate the data points into classes with the help of a kernel function (Singh, 2019). This method can uniquely overcome complex non-linear relationships with the help of the kernel function, which maps the input data into high-dimensional feature space (Hou et al., 2023). In this study, the SVM optimization was achieved by selecting the Gaussian radial basis function as the designated kernel for operations, with a chosen C value of 0.2 assigned to control the smoothness of the decision boundaries.

3.2.2.3 Decision Tree (DT)

Decision tree classifiers are algorithms that use the rules to make decisions by recursively partitioning the input space into subsets based on the most significant features. These classifiers are easy to interpret and thus valuable in decision-making and classification problems. The stabilized and most accurate decision tree chosen for the study has configured the Splitter parameter 'best' and set the Min Impurity Decrease to 0.1.

3.2.2.4 k-Nearest Neighbors (k-NN)

k-Nearest Neighbors is an analytical ML method used for classification and regression-based problems. This method calculates the distance from each class to the objects, and the class with the minimum distance is assigned as the object's class

(Singh, 2019). The classifier's accuracy was improved by fine-tuning parameters, including the number of neighbors, search method (grid and random), and distance metric. In the present study, the model's performance was optimized by employing the Euclidean distance, 5 neighbors, and a random search method.

3.2.2.5 Multi-Layer Perceptron (MLP)

Multi-layer Perceptron is a widely used artificial neural network for classification, data mining, and modeling processes (Asadi & Pourhossein, 2021). The structure of an MLP model comprises an input layer, an output layer, and one or more intermediate layers dedicated to receiving decision factors, outputting classified results, and performing the computational process of the model, respectively (Sachit et al., 2022a). In supervised classification, the MLP model is trained through backpropagation, wherein the model's accuracy is improved in each iterative process by adjusting the model with a specified learning rate.

The MLP model utilized in this study employed a Multi-layer Perceptron (MLP) model with 4 hidden layers, comprising 100, 80, 60, and 40 Perceptron, respectively. The chosen activation function was the rectifier linear unit, and the model underwent training with a maximum of 5000 iterations. An early stopping method was implemented to prevent model overfitting, which monitored the validation fraction set at 0.2.

3.2.3 Model evaluation, validation, and selection

Performance evaluation is an essential task in ML projects, which allows the assessment of the accuracy of predictions using a new data set (Hou et al., 2023). After implementing the methods using Python’s executable code, the performance of each method was assessed by utilizing confusion matrices to derive performance metrics, including Accuracy, Precision, F1 Score, and Receiver Operating Characteristic (ROC) curves, which calculate the area under the curves. Finally, the model that achieved the highest F1 score and the most significant area under the curve was chosen as the optimal one to advance the research.

3.2.3.1 Receiver Operating Characteristic (ROC) curves

The ROC curves are often used to compare the performance of classification tasks by illustrating the actual positive rate of predictions against the false positive rate (Sachit et al., 2022a). The area under the curve is the quantitative measure that evaluates the quality of the ROC curves. The AUC values can range from 0.5 to 1, 1 indicating a perfect model. Since the study dataset contained five suitability classes as labels, separate receiver operating characteristics were recorded for each class to calculate the areas under the curves.

3.2.3.2 Accuracy

The measuring accuracy quantifies the performance of the ML models by obtaining the ratio of correctly classified pixels (true positives) to the total number of pixels, which is the sum of true positives, true negatives, false negatives, and false positives (Itza Alejandra et al., 2020). The accuracy of five ML models was independently computed for five suitability classes, utilizing the developed confusion matrix. Then, the final accuracy value was derived by calculating the unweighted average across the different land use classes.

$$Accuracy = \frac{TP}{TP+TN+FP+FN} \qquad (3.2)$$

3.2.3.3 Precision

Precision is the ratio between true positives and all positives captured by the model, namely true and false positives (Sirefelt Rickard, 2004). It was employed in the study to measure the ML model's capacity to predict positive instances accurately. Similar to the accuracy calculations, the precision of independent classes was calculated to obtain the average precision for each ML model.

$$Precision = \frac{TP}{TP+FP} \qquad (3.3)$$

3.2.3.4 F1 score

The F1 Score is a matrix used to measure a model's overall performance by combining precision and recall into a single matrix (Shrestha & Vanneschi, 2018). It measures the model's ability to capture all positive instances by getting the harmonic mean of precision and recall, where precision measures the ratio between true positives and the sum of true and false positives. In contrast, recall measures the ratio between true positives and the sum of true positives and false negatives (Abdollahi et al., 2020). Given the multiclass classification nature, this study employs the macro-average F1 score for the assessment.

$$F1\ Score = \frac{2*precision*recall}{precision+recall} \quad (3.4)$$

3.2.4 Explainable AI

Explainable AI (XAI) is a concept aiming to explain the internal process of an ML model by providing user-understandable explanations for the method, the process, and the output of an ML model (Sachit et al., 2022a). In this study, the popular XAI algorithm, known as SHAP, is used to explain the RF model, which was identified as the most accurate among the five selected models. Shapely introduced the SHAP algorithm (Vajda et al., 1951) in 1951 to evaluate the contribution of individual players to a n-persons game where more than one player is contributing to the game. It was subsequently expanded in artificial intelligence to interpret model predictions in 2017 by Scott Lundberg (Lundberg & Lee, 2017). The SHAP algorithm is applied in various forms, such as TreeSHAP, DeepSHAP, and Kernel SHAP. This study uses TreeSHAP, designed explicitly for ML models based on tree structures (Sachit et al., 2022b).

The SHAP algorithm has two explainability techniques: global and local explanations, where the global explanations describe the model in general. In contrast, local explanations describe every observation in the model (Dallanoce, 2022). In this study, the global interpretations of the model were obtained by illustrating the summary plots of the model. These explanations aim to map the contribution of each independent variable in assessing the locational suitability of solar power plants. Then, the local explanations were obtained through force plots, highlighting the factors that significantly contributed to the prediction by identifying the effect of each independent variable on classifying an individual pixel in the suitability map.

3.2.5 Solar power plant suitability map

As the next step of the study, a suitability map for establishing solar power plants in the United States of America was generated using the trained model. After removing the protected areas, the model's predictions were displayed cartographically as a map, serving as a valuable resource for the decision-making process in the future. The quantitative analysis of the spatial distribution was subsequently carried out to derive insights regarding the percentages of land parcels and areas suitable for mapping solar power plants. In the final stage of the methodology, a comparative analysis was conducted between the existing solar power plants classified according to their efficiency values, and those are obtained through model predictions. This assessment aimed to identify discrepancies between the ground truth data and the model predictions.

4. RESULTS AND DISCUSSION

This chapter focuses on the visualization and discussion of the outcomes derived from the conducted study. The initial sections showcase the outcomes of the pre-processing steps. The following section illustrates the results obtained during the model selection and accuracy assessment. The final two sections provide a detailed account of the results obtained for explainable AI and the final solar suitability maps.

4.1 Results

4.1.1 Data pre-processing

After applying various data pre-processing steps, including reprojection, clip, raster-to-vector conversion, and resampling, Figure 4.1 – 4.12 presents the final raster layers of the twelve independent variables.

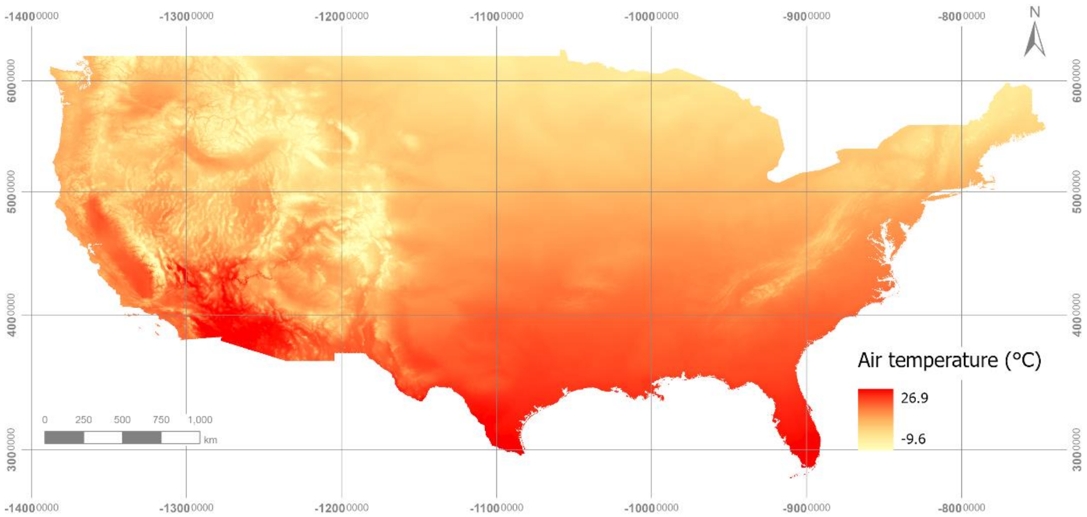

Figure 4.1: Air temperature

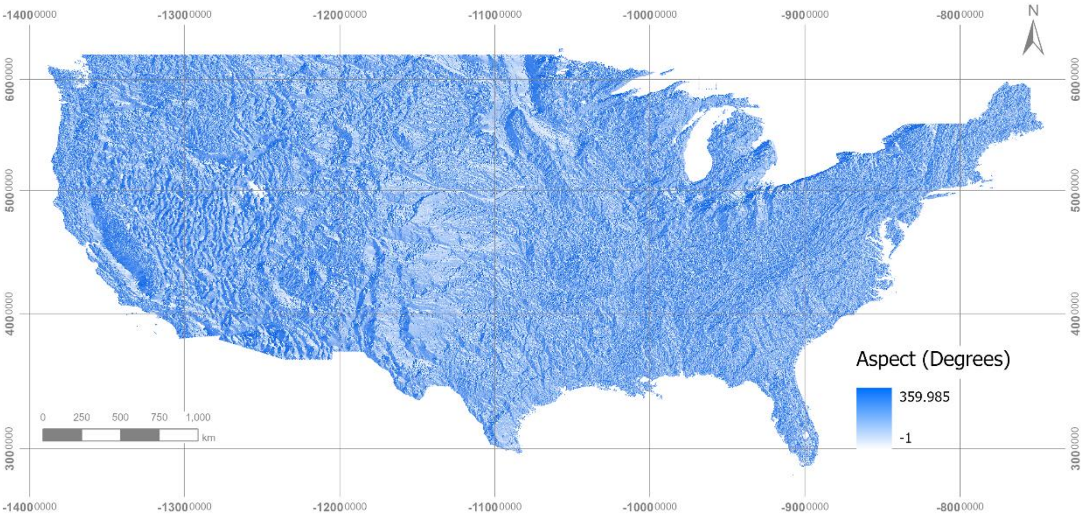

Figure 4.4: Aspect

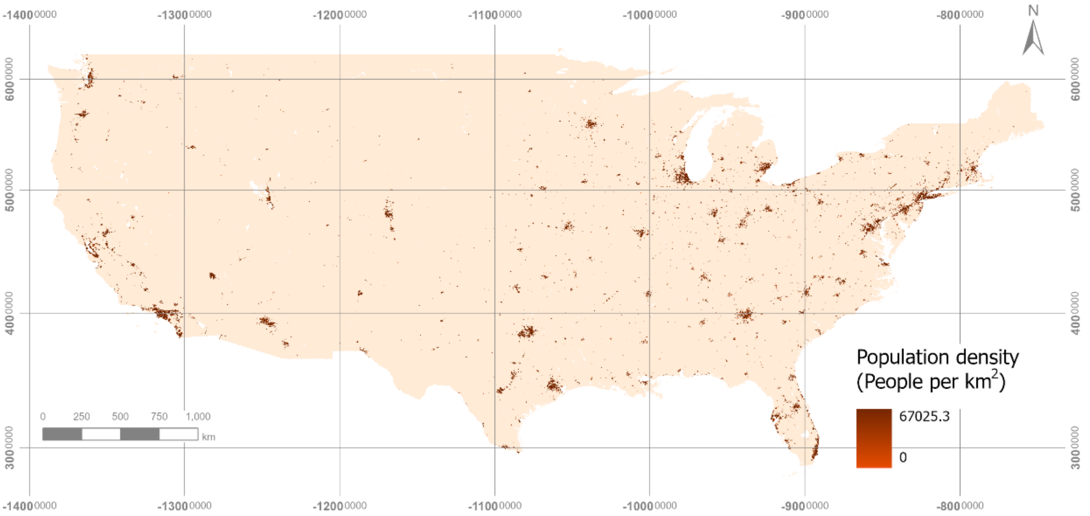

Figure 4.3: Population density

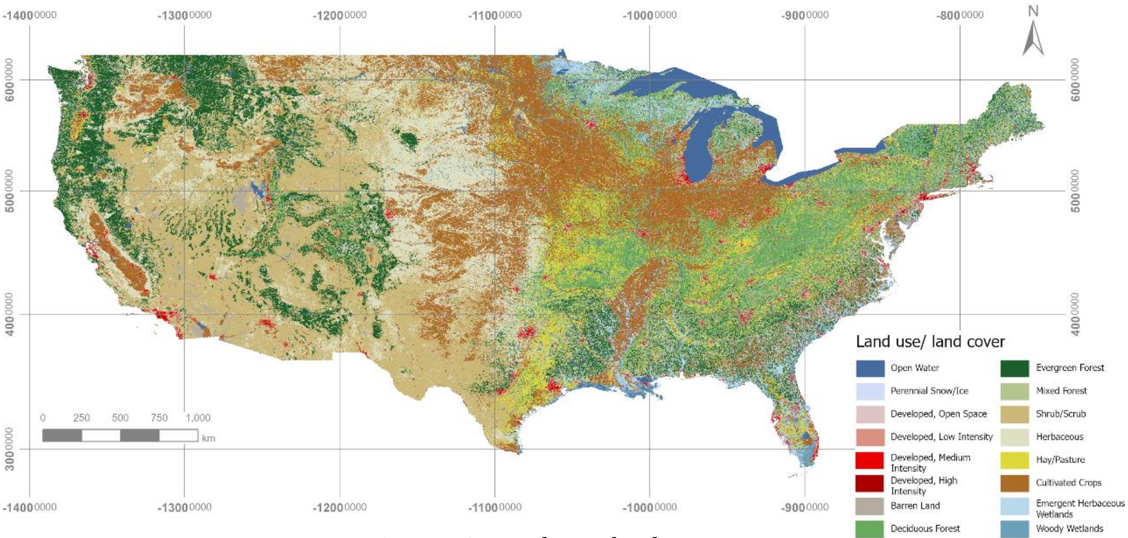

Figure 4.2: Land use / land cover

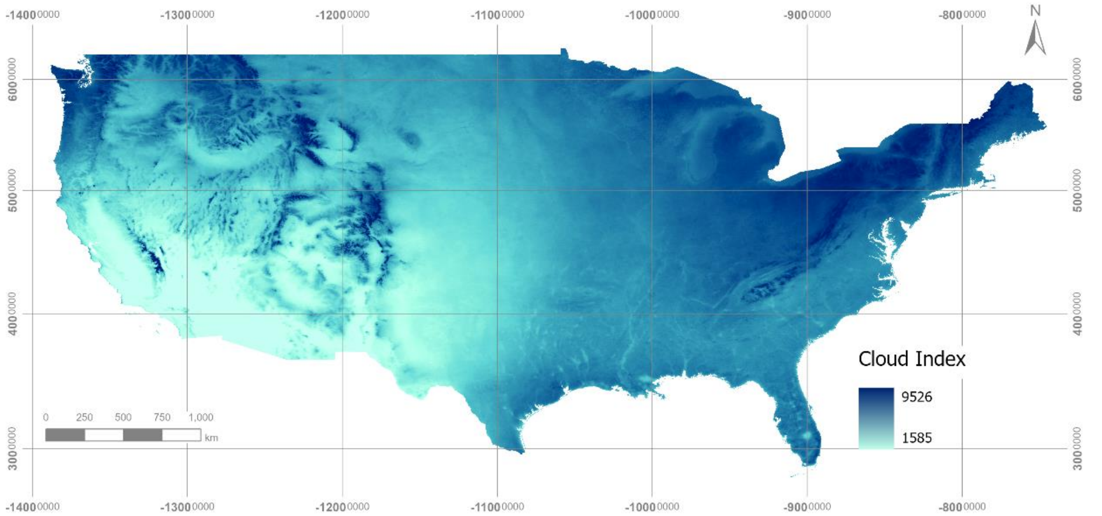

Figure 4.7: Cloud index

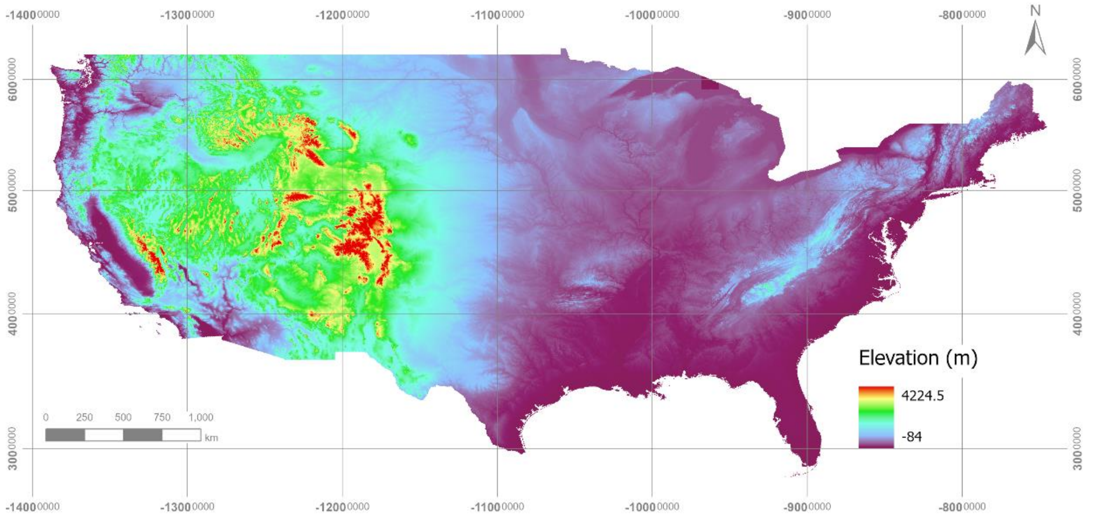

Figure 4.6: Elevation

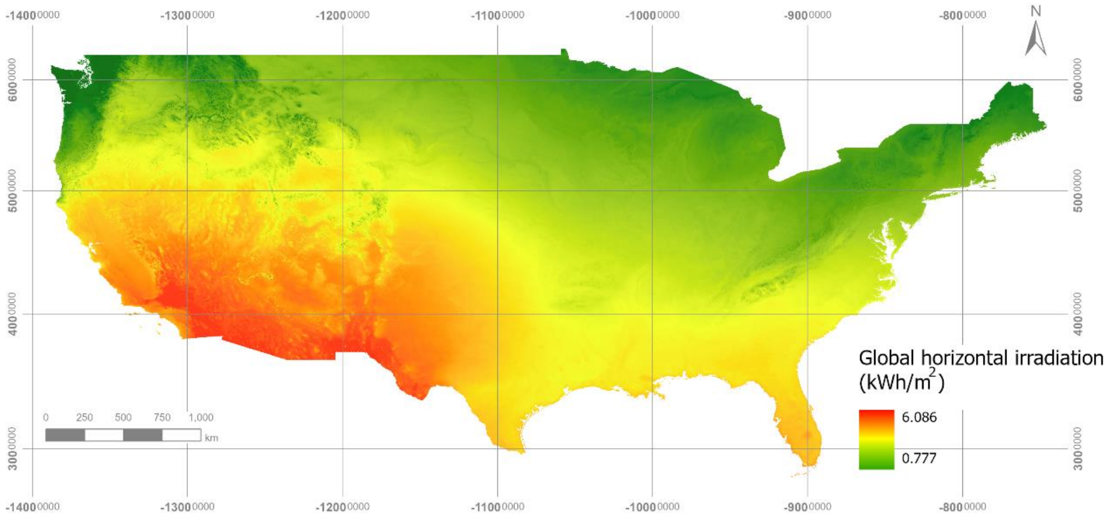

Figure 4.5: Global horizontal irradiation

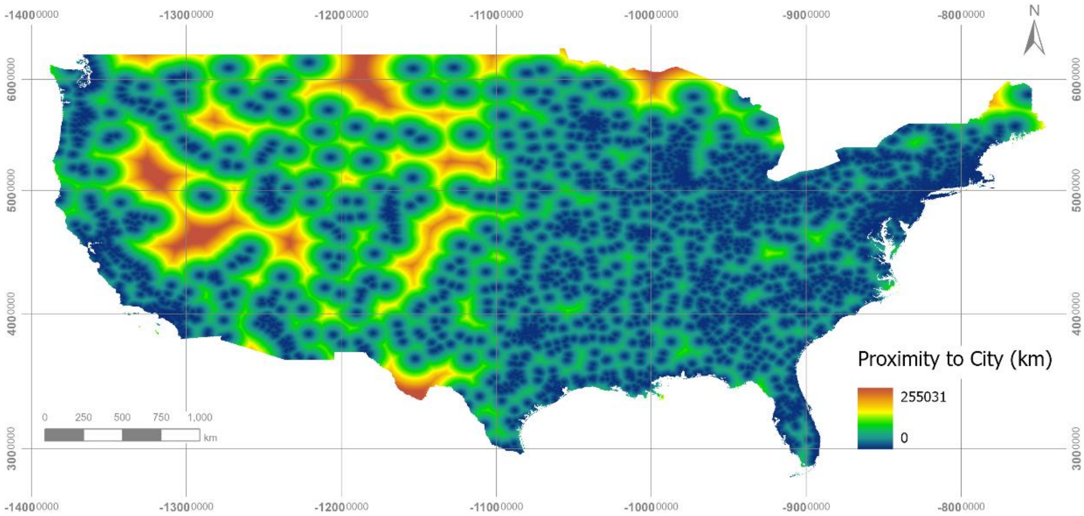

Figure 4.10: Proximity to city

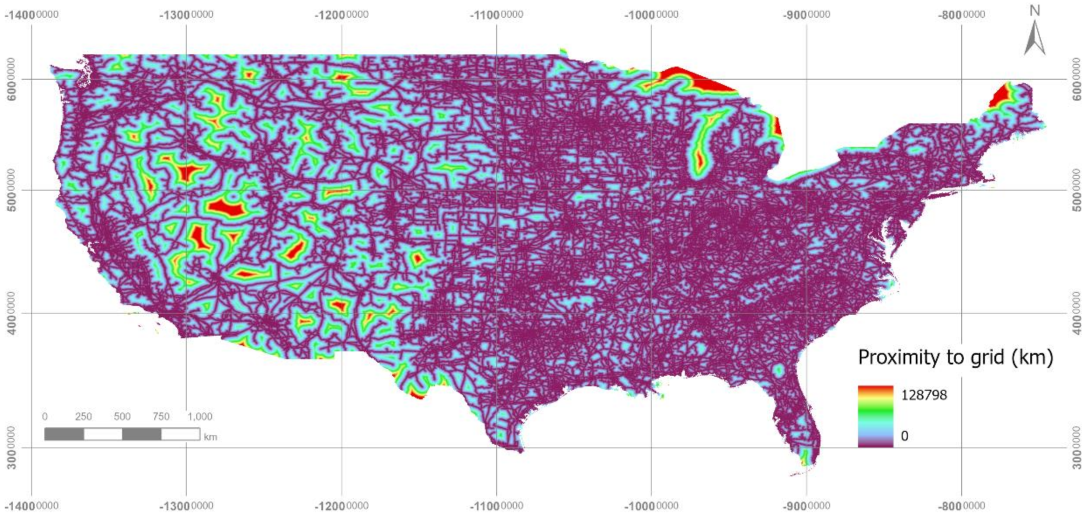

Figure 4.9: Proximity to road

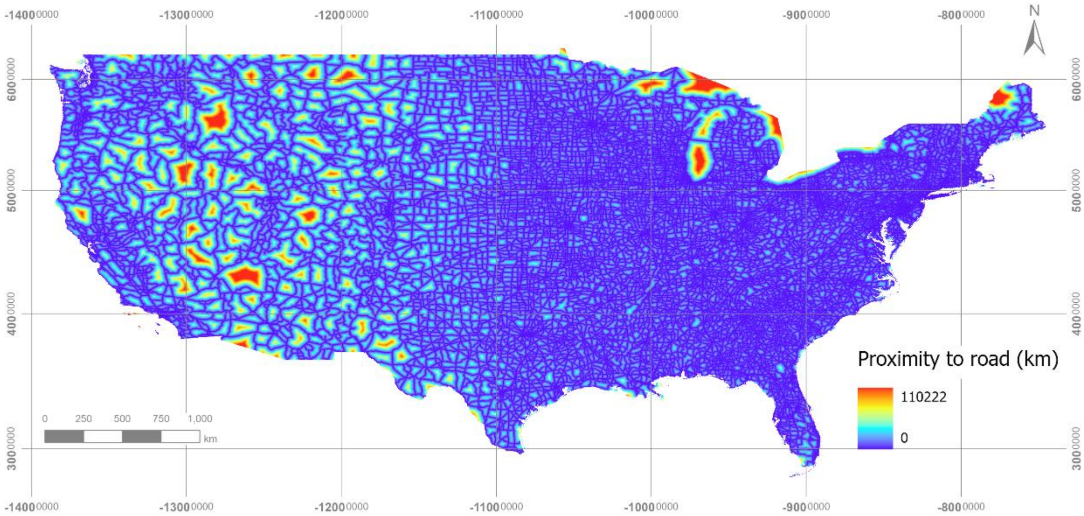

Figure 4.8: Proximity to grid

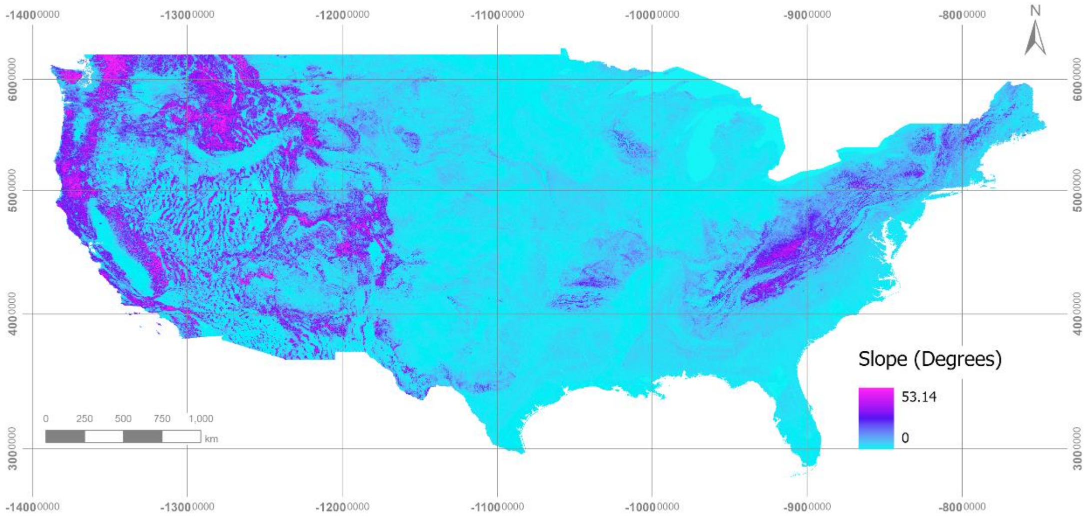

Figure 4.12: Slope

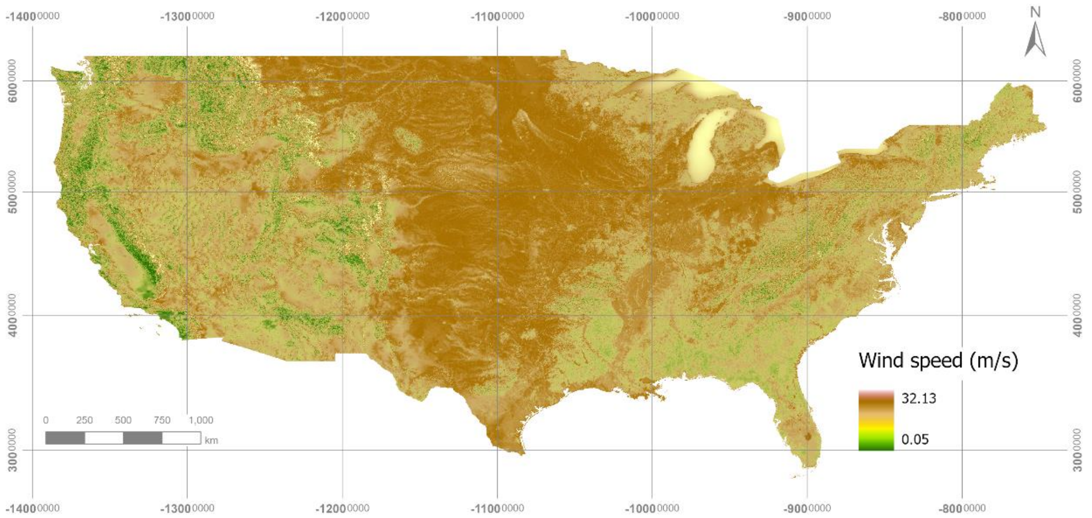

Figure 4.11: Wind speed

These twelve thematic layers were subsequently utilized to extract the data at the solar power plant locations of the United states of America. The resulting geodatabase consisted of 4940 observations to be utilized for the ML models.

4.1.2 Multicollinearity and Collinearity assessment

The results of the multicollinearity assessment are listed in Table 4.1. Given that the multicollinear variables exhibit Tolerance(Ti) values below 0.1 and VIF values higher than 10, it can be concluded that there is no multicollinearity among the variables to be considered for subsequent steps. The outcomes of the collinearity assessment are presented in Table 4.2. A positive correlation was observed between the solar radiation and the air temperature of the area, while a negative correlation was identified between the cloud index and the solar radiation. Consequently, the ML models were trained with and without considering the collinear variables.

Conditioning factors	**Tolerance (Ti)**	**VIF**
Air temperature	0.943405	1.059990
Aspect	0.995113	1.004911
Proximity to the city center	0.964059	1.037281
Cloud Index	0.841803	1.187926
Elevation	0.938395	1.065650
Solar radiation	0.834098	1.198900
Proximity to grid line	0.999143	1.000857
Land cover	0.999979	1.000021
Population density	0.984489	1.015755
Proximity to road network	0.965136	1.036124
Slope	0.987504	1.012655
Windspeed	0.999452	1.000549

Table 4.1: Multicollinearity statistics for the conditioning factors

AT	1.00	-0.03	0.14	-0.56	-0.18	0.77	-0.12	-0.01	0.01	0.25	-0.26	0.28
A	-0.03	1.00	-0.09	0.06	-0.07	-0.08	0.00	0.01	0.07	-0.05	0.02	-0.06
PC	0.14	-0.09	1.00	-0.14	0.31	0.22	0.01	0.01	-0.18	0.22	-0.09	0.09
CI	-0.56	0.06	-0.14	1.00	-0.39	-0.91	0.06	0.16	-0.05	-0.35	0.20	0.24
E	-0.18	-0.07	0.31	-0.39	1.00	0.36	0.07	-0.05	-0.10	0.22	0.05	0.07
SR	0.77	-0.08	0.22	-0.91	0.36	1.00	-0.07	-0.13	0.00	0.37	0.22	-0.24
PD	-0.12	0.00	0.01	0.06	0.07	-0.07	1.00	0.02	-0.03	0.07	0.07	0.03
LULC	-0.01	0.01	0.01	0.16	-0.05	-0.13	0.02	1.00	-0.20	0.01	-0.07	0.20
PR	0.01	0.07	-0.18	-0.05	-0.10	0.00	-0.03	-0.20	1.00	-0.11	0.02	-0.23
PG	0.25	-0.05	0.22	-0.35	0.22	0.37	0.07	0.01	-0.11	1.00	-0.07	0.03
S	-0.26	0.02	-0.09	0.20	0.05	0.22	0.07	-0.07	0.02	-0.07	1.00	-0.13
WS	0.28	-0.06	0.09	0.24	0.07	-0.24	0.03	0.20	-0.23	0.03	-0.13	1.00
	AT	**A**	**PC**	**CI**	**E**	**SR**	**PG**	**LULC**	**PD**	**PR**	**S**	**WS**

Table 4.2: Collinearity statistics for the conditioning factors

4.1.3 Machine learning model selection

The accuracies achieved for the five ML algorithms are presented in Table 4.3, considering both scenarios: with the removal of collinear conditioning variables and without their removal. The results indicate higher values when the ML models are trained without eliminating the collinear variables. Notably, the Random Forest classifier emerged as the most effective classifier for mapping the solar power plant suitability classes, closely followed by the Decision trees and K-nearest neighbor classifier. Support Vector Machines' classifier achieved the lowest accuracy across all accuracy measures.

Model name	Without eliminating the collinear variables			After eliminating the collineated variables		
	F1 score	Accuracy	Precision	F1 score	Accuracy	Precision
Random Forest	0.881	0.881	0.881	0.854	0.855	0.852
Support vector machines	0.395	0.424	0.402	0.376	0.378	0.377
K-nearest neighbor	0.742	0.751	0.747	0.706	0.716	0.712
Decision tree	0.758	0.758	0.757	0.736	0.737	0.735
Multi-layer Perceptron	0.632	0.637	0.634	0.607	0.612	0.609

Table 4.3: Accuracies achieved for the five ML algorithms

In addition, the calculated values of the Areas Under the Curves (AUC) provided confirmation of the Random Forest (RF) model's suitability for the selected study, as it achieved the highest values, approaching 1, for all suitability classes. The receiver operating characteristics (ROC) curves, utilized for calculating the AUC values, can be found in the document's Annexes A. Finally, the Random Forest model was chosen to implement explainable AI algorithms further to describe the results of the model's outcome.

4.1.4 Explainable AI

4.1.4.1 Summary plot (Global explanations)

The global explanations obtained by the SHAP algorithm are illustrated in Figure 4.13, which describes the average importance of each conditioning factor for the prediction outputs derived from the Random Forest model. The Y axis of the plot lists the variables in descending order of their importance to the model. Solar radiation appeared as the highly impacted variable for the predicted outputs, followed by the cloud index and air temperature. On the other hand, the aspect of the area appeared to be the least impactful factor in the predictions.

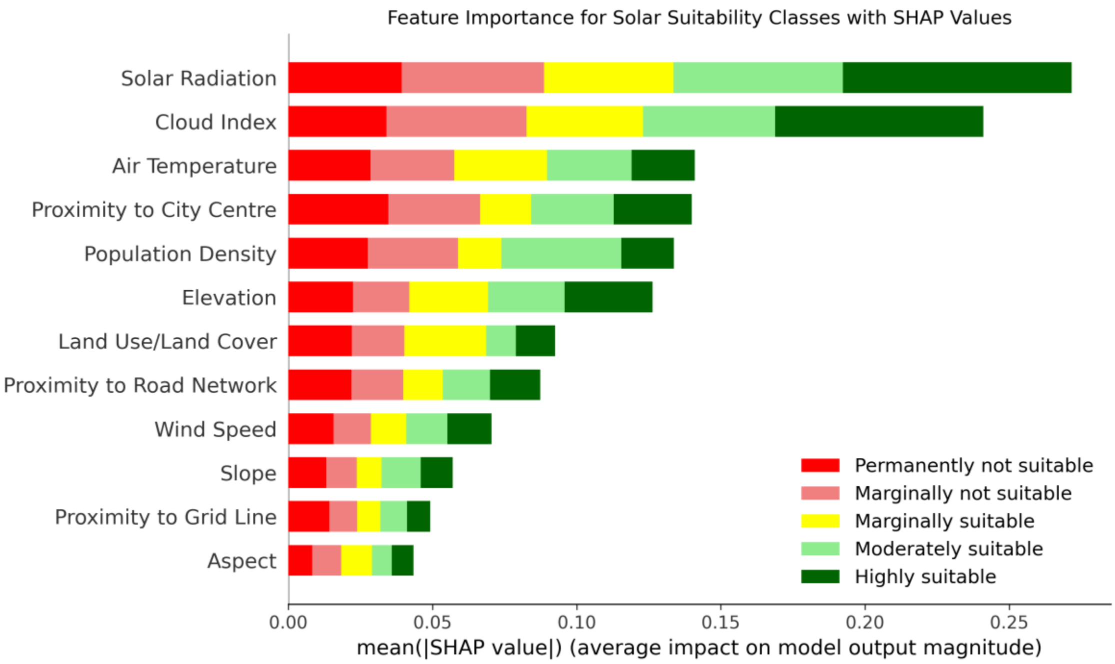

Figure 4.13: Relative average impact of conditioning factors

4.1.4.2 Force plot (Local explanations)

In addition, the significance of conditioning factors for the five suitability classes was computed individually, as demonstrated in Figure 4.14 for the highly suitable class. The individual plots for the other classes can be found in the document's Annexes B. Bluish dots of these plots signify lower values of the factor, while Reddish dots represent higher values. These individual visualizations offer insights into the significance of conditioning factors for each class and clarify how a factor's low and high values impact the model predictions, positively or negatively. For instance, as depicted in Figure 4.14, solar radiation emerged as the most influential variable for predicting the highly suitable class. Additionally, the visual representation indicates that lower values of the solar radiation variable have a negative impact on the predictions, while higher values have a positive impact. This implies that for a site to be well-suited for establishing a solar power plant, it needs to receive increased insolation.

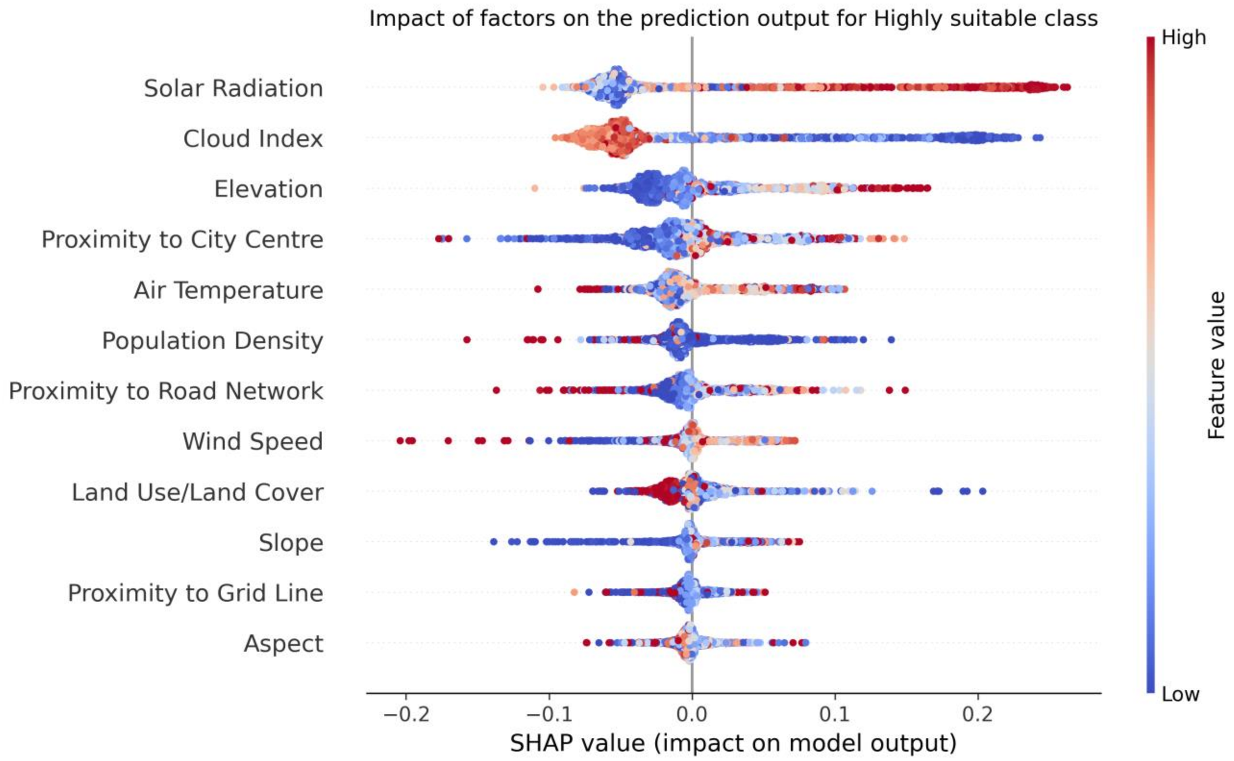

Figure 4.14: Significance of conditioning factors for the predictions of highly suitable class

4.1.5 Weights of the conditioning factors

Determining the relative weight for each conditioning factor in the ML models involves calculating the mean decrease in impurity (MDI), also known as "Gini importance. (Sachit et al., 2022). The calculated relative importance of the final Random Forest model is illustrated in Figure 4.15; the numerical representation of these weights allows for a detailed examination of each independent factor's impact on the predictive capabilities of the trained Random Forest model.

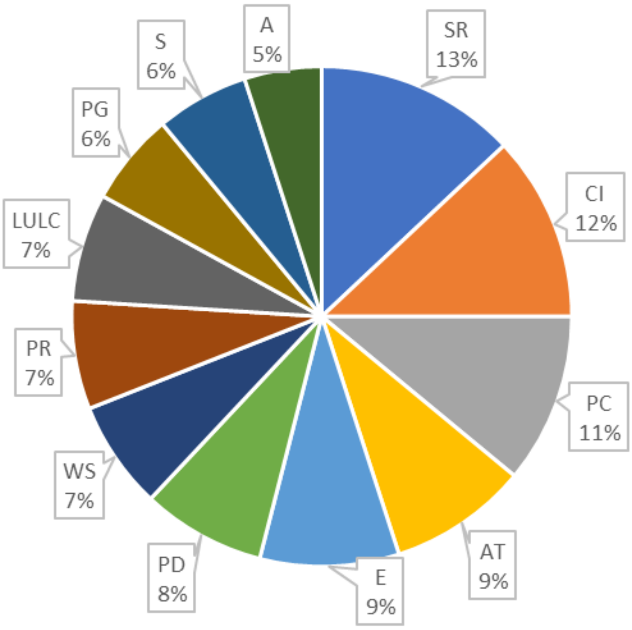

Figure 4.15: Weights of the factors

The outcomes reveal that solar radiation is the primary determinant in solar power plant suitability modeling, contributing 13% to the model. Following closely, the cloud index and proximity to the city center emerge as the second and third most significant factors, with influences of 12% and 11%, respectively. Subsequently, air temperature and elevation share equal importance at 9% each in the model. The least impactful factor in the model is the aspect, contributing only 5% to the overall importance of the RF model.

4.1.6 Solar power plant suitability map

Figure 4.16 depicts the resulting map indicating the suitability of the solar power plant generated through the Random Forest classification model. Before creating the map, the predictions of the ML model underwent a filtering process, which was used to assign the "Permanently not suitable" class to protected areas.

The quantitative examination of the ultimate suitability map is presented in Table 4.4, indicating that around 5% of the total land area in the United States is highly suitable for constructing solar power plants, ensuring optimal operational efficiency. Additionally, 55% of the land is moderately suitable for such establishments. Conversely, approximately 9.5% of the total land area, equivalent to approximately

766654 km^2, is deemed not permanently suitable for solar power plant construction. This limitation is primarily attributed to environmentally protected areas reserved throughout the country.

Class	Area (sq. km)	Percentage (%)
Highly suitable	389666	4.9
Moderately suitable	4377279	55.07
Marginally suitable	2329314	29.3
Marginally not suitable	86039	1.08
Permanently not suitable	766654	9.65

Table 4.4: Quantitative examination of the final suitability map

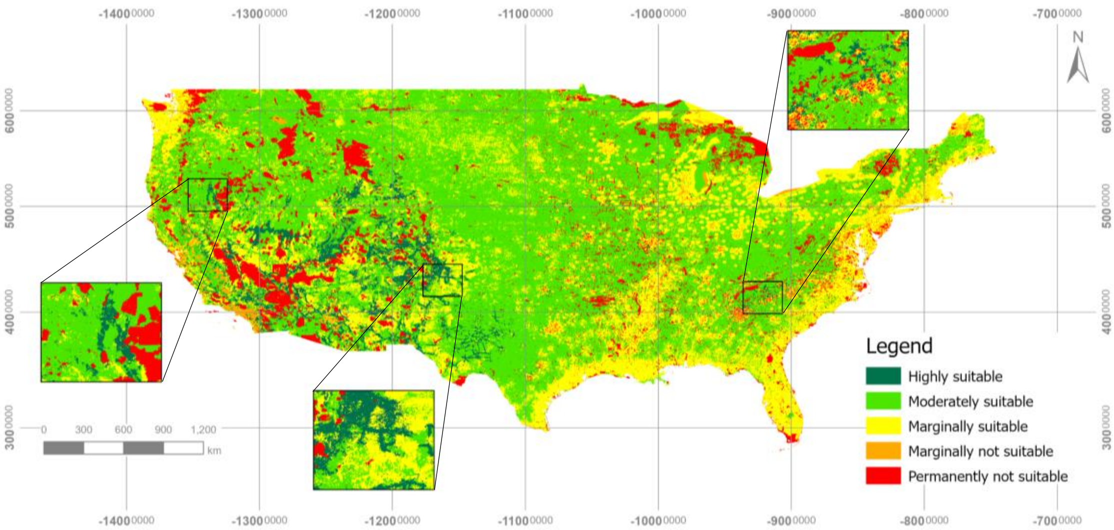

Figure 4.16: Final suitability map

4.1.7 Comparative analysis of ground truth data and model predictions

Figure 4.17 illustrates the overlay of existing power plants over the suitability map produced using the Random Forest model. The quantitative analysis conducted over the overlaying map is presented in , displaying the number of existing solar power plants within each suitability class of the map along with the actual number of the solar power plants located in the respective class that was calculated using the method described in the section 3.2.1.

No:	Class	Ground truth data	Prediction Map	Difference	Model Accuracy
1	Highly suitable	173	187	14	Overestimate
2	Moderately suitable	609	383	26	Underestimate
3	Marginally suitable	2996	3006	10	Overestimate
4	Marginally not suitable	939	1114	25	Underestimate
5	Permanently not suitable	223	250	27	Overestimate

Table 4.5: The distribution of predicted and existing solar power plants over the study region

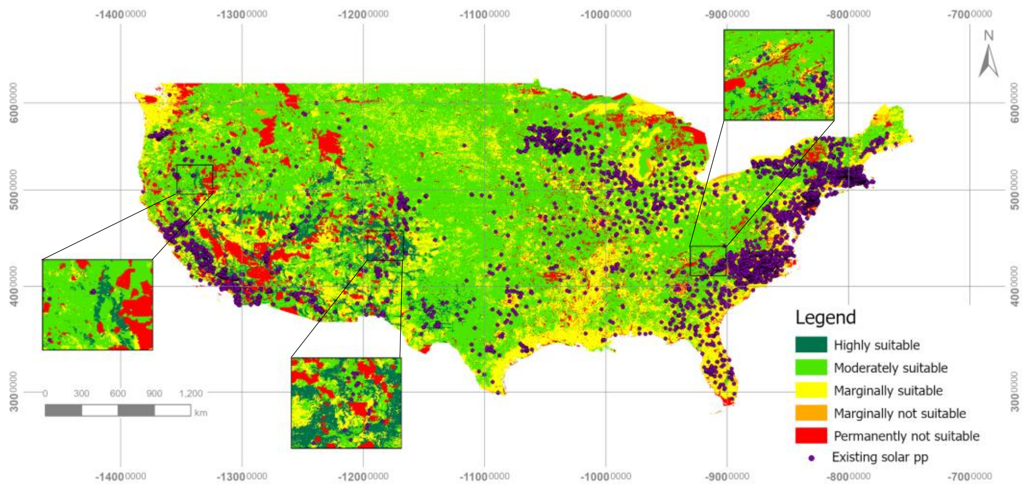

Figure 4.17: Existing solar power plants over the suitability map produced with the RF model

The observed discrepancies suggest that the predictions overestimate three classes and underestimate two classes. However, given the slight differences in values, it can be assumed that the model performed well in spatial distribution.

4.2 Discussion

The conventional approaches to site suitability assessment in scientific research often involve using multicriteria-based analysis methods to identify the optimal locations for installing solar power plants. However, these conventional methods have inherent limitations related to the subjective selection of criteria and the categorization of influencing factors. Such subjectivity can introduce biases and pose limitations in real-world renewable energy scenarios, especially when there is insufficient knowledge or maturity in the methodology. Nevertheless, the proposed study uses advancements in the availability of real-world digitally formed spatial data, geographical information systems, and remotely sensed conditional factors to develop more mature and less biased ML-based solutions that effectively address locational suitability issues. The performance measures of the current study, where three out of the five selected models achieved an accuracy exceeding 74%, affirm the suitability and effectiveness of such methods in decision-making projects.

The highest accuracy, precision, F1 score, and AUC were achieved using the Random Forest classifier, followed by the decision trees, K-nearest neighbors, and Multi-layer Perceptron in the given study. These results align with previous findings, emphasizing the efficiency and accuracy of the Random Forest model for locational suitability research (Sachit et al., 2022), (Aamir Shahab & M.P. Singh, 2022), (Sun et al., 2023).

Moreover, the results of the importance of conditional factors calculated with the Random Forest model reveal that the solar radiation, cloud index, and proximity to the city center of the research area employ the most significant influences on the selections. These weights offer valuable insights for choosing a location to host a solar power plant in an economically efficient environment. The factors pinpointed as the most influential by the model are consistent with the recommendations derived from numerous studies in the field of locational choice research (Akkas et al., 2017), (Yousefi et al., 2018), (Gherboudj & Ghedira, 2016), (Husein et al., 2023).

Meanwhile, the outcomes derived through the XAI approach have contributed to a deeper comprehension of the selected model and the conditioning factors that significantly affect a specific location's classification within a particular suitability class. The global explanations of the Random Forest (RF) model offered insights into the mean influence of each prediction across five suitability classes. Solar radiation exhibited the highest mean SHAP value, followed by cloud index and air temperature.

In addition, the local explanations elucidated the impact of each factor on the model concerning their assignment to specific categories. Solar radiation emerged as the most influential factor for the highly suitable class, with lower values negatively impacting the class and higher values positively contributing to the class. The analysis offered a thorough overview of each factor, illustrating how lower and higher values impact each classification class. The proposed study concludes that the ML models are suitable and that Explainable AI approaches effectively explain the selected ML-based model for solar power plant suitability mapping at a resolution of 1 km^2 for the continental USA. Further, it is essential to acknowledge that the current study has primarily focused on evaluating the physical, climatic, and construction cost-related parameters favorable to the selection process of solar power plant sites. However, for a comprehensive assessment, it is necessary to also account for additional economic and political factors.

5. CONCLUSIONS AND RECOMMENDATION

This chapter summarizes the study's conclusion in alignment with the predefined research objectives. Additionally, it addresses the identified limitations and recommendations for future research endeavors, aiming to enhance and refine the study's outcomes.

5.1 Conclusions

The study's primary objective aimed to develop an explainable AI-based ML model for mapping the suitability of solar power plant locations by considering the efficiency of the plant as the dependent variable of the ML models instead of considering all existing plants in the ideal locations. Incorporating efficiency into the model involved successful implementation, where calculations were executed using energy production and capacity data from the U.S. Energy Information Administration (EIA). However, the absence of cutoff values for classifying plants into suitability classes led to developing a hypothetical assumption, elaborated in Chapter 3, section 3.2.1.

Five ML algorithms (Random Forest, Decision tree, K-nearest neighbors, multi-layer Perceptron, and Support vector machines) were employed in the study using real-world data. Results confirmed the superiority of the Random Forest model for the study with high-performance values for accuracy metrics: F1 Score: 0.881, Precision: 0.881, Accuracy: 0.881. Consequently, the Random Forest algorithm can be suggested as a suitable model for the suitability assessment frameworks of solar power plants. The conclusive suitability map developed utilizing a chosen Random Forest indicates that approximately 4.90% of the total land area in the USA is recommended for constructing solar power plants to achieve higher efficiency in energy production. Furthermore, 55.07% of the land is classified as moderately suitable, with 29.30% and 1.08% identified as marginally suitable and marginally not suitable, respectively.

These findings provide a practical and actionable overview for decision-makers and stakeholders in the solar energy sector, offering insights into the most suitable areas for optimal and efficient solar power plant construction across the country. The conclusions drawn from addressing the established research questions in pursuit of the primary objective are as follows:

1. **What are the key factors that significantly influence the suitability of solar power plant site selection process?**

   In conclusion, examining key factors influencing the suitability of solar power plant site selection involved an extensive literature review, encompassing studies beyond and within the study area. This review identified 14 factors that significantly influence the site selection process as Solar radiation (SR), Cloud index (CI), Proximity to city centers (PC), Population density (PD), Elevation (E), Proximity to road network (PR), Land use and land cover (LULC), Proximity to grid line (PG), Air temperature (AT), Wind speed (WS), Aspect (A), Slope (S), Humidity (HD), and Natural disasters (ND). However, although humidity and occurrences of natural disasters in the area were identified as significant, those were excluded from the study due to the data unavailability. This highlights the importance of the identified factors and underscores the need for comprehensive data availability for a more exhaustive analysis in future studies.

2. **Which ML-based classification method is the most accurate for creating a suitability map to identify potential solar power sites using existing solar productivity data?**

   In conclusion, the investigation into ML-based classification methods aimed at creating a suitability map for identifying potential solar power sites yielded valuable insights. The study employed five distinct methods: Random Forest, Decision trees, K-nearest neighbors, Multi-Layer Perceptron, and Support Vector machines. Through rigorous evaluation, Random Forest emerged as the most accurate classification method, demonstrating superior performance in terms of accuracy. The selection of Random Forest as the preferred model for generating a suitability map underscores its efficacy in leveraging existing solar productivity data for precise identification of potential solar power sites. This conclusion provides practical guidance for future endeavors in solar site selection, emphasizing the importance of employing Random forests for optimal accuracy and reliability.

3. **What is the influence of each factor on the final suitability map?**

   In conclusion, the SHAP analysis, conducted through explainable AI techniques, highlighted the critical role of individual environmental, geomorphological, spatial, and climatic criteria in the spatial decision-making process for predicting the suitability class of solar power plant locations. These factors were ranked in order of importance, from highest to lowest, as follows: SR - 13%, CI - 12%, PC - 11%, AT - 9%, E - 9%, PD - 8%, WS - 7%, PR - 7%, LULC - 7%, PG - 6%, S - 6%, and A - 5%

5.2 Limitations and Recommendations

While this study contributes valuable insights into Explainable ML-based solar power plant selection methodologies, it is essential to acknowledge certain limitations inherent in the methodology and data sources. These limitations may impact the generalizability and applicability of the findings.

1. **Unavailability of data**:

   Limited data availability for humidity and the occurrence of natural disasters restricted their inclusion in the analysis, emphasizing the need for comprehensive datasets in future studies.
2. **Limited literature on efficiency-based classification:**

   Scarce literature on classifying solar power plants based on efficiency constrained the study's depth, highlighting the importance of expanding research in this area.
3. **Insufficient recorded solar power plants:**

   The study faced limitations due to insufficient solar power plants with recorded production and consumption data, particularly for large-scale site suitability mapping. Collaborative efforts are needed to compile a more extensive dataset for future research.

For future recommendations, the study would underscore the importance of incorporating solar panels' tilt and sun orientation as influential factors, thus enhancing the comprehensive understanding of site suitability for solar power plants.

6. BIBLIOGRAPHIC REFERENCES

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2020). Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sensing*, *12*(9). https://doi.org/10.3390/RS12091444

Akinsola, J. E. T., Awodele, O., Kuyoro, S. O., & Kasali, F. A. (2019). Performance Evaluation of Supervised Machine Learning Algorithms Using Multi-Criteria Decision Making Techniques. International Conference on Information Technology in Education and Development (ITED), Mcdm, 17–34. https://www.academiainformationtechnology.org/ited2019/uploads/8135_File_03ITED19041 IEEE Paper Format Performance Evaluation of Supervised Machine Learning Algorithms Using MCDM Techniques NEW (1).pdf

Akkas, O. P., Erten, M. Y., Cam, E., & Inanc, N. (2017). Optimal Site Selection for a Solar Power Plant in the Central Anatolian Region of Turkey. *International Journal of Photoenergy*, *2017*(7 June 2017). https://doi.org/10.1155/2017/7452715

Al-Ruzouq, R., Shanableh, A., Yilmaz, A. G., Idris, A. E., Mukherjee, S., Khalil, M. A., & Gibril, M. B. A. (2019). Dam site suitability mapping and analysis using an integrated GIS and machine learning approach. *Water (Switzerland)*, *11*(9). https://doi.org/10.3390/w11091880

Alami Merrouni, A., Elwali Elalaoui, F., Mezrhab, A., Mezrhab, A., & Ghennioui, A. (2018). Large scale PV sites selection by combining GIS and Analytical Hierarchy Process. Case study: Eastern Morocco. *Renewable Energy*, *119,* 863–873. https://doi.org/10.1016/j.renene.2017.10.044

Ali, R., Hussain, A., Nazir, S., Khan, S., & Khan, H. U. (2023). Intelligent Decision Support Systems—An Analysis of Machine Learning and Multicriteria Decision-Making Methods. Applied Sciences, 13(22), 12426. https://doi.org/10.3390/app132212426

Almansi, K. Y., Shariff, A. R. M., Abdullah, A. F., & Ismail, S. N. S. (2021). Hospital site suitability assessment using three machine learning approaches: Evidence from the gaza strip in Palestine. *Applied Sciences (Switzerland)*, *11*(22), 1–22. https://doi.org/10.3390/app112211054

Asadi, M., & Pourhossein, K. (2021). Neural network-based modelling of wind/solar farm siting: a case study of East-Azerbaijan. *International Journal of Sustainable Energy*, *40*(7), 616–637. https://doi.org/10.1080/14786451.2020.1833881

Asadi, M., Pourhossein, K., Noorollahi, Y., Marzband, M., & Iglesias, G. (2023). A New Decision Framework for Hybrid Solar and Wind Power Plant Site Selection Using Linear Regression Modeling Based on GIS-AHP. *Sustainability (Switzerland)*, *15*(10). https://doi.org/10.3390/su15108359

Bayounis, A., & Eldamaty, T. (2022). Applying Geographic Information System (GIS) for Solar Power Plants Site Selection Support in Makkah. *Technology and Investment*, *13*(02), 37–58. https://doi.org/10.4236/ti.2022.132003

Dallanoce, F. (2022). *Explainable AI: A Comprehensive Review of the Main Methods*. Www.Medium.Com. https://medium.com/mlearning-ai/explainable-ai-a-complete-summary-of-the-main-methods-a28f9ab132f7

EIA, U. states of A. (2023). *U.S. energy facts explained*. U.S. Energy Information Administration. https://www.eia.gov/energyexplained/us-energy-facts/#:~:text=U.S. total annual energy production,primary energy production in 2022.

Gherboudj, I., & Ghedira, H. (2016). Assessment of solar energy potential over the United Arab Emirates using remote sensing and weather forecast data. *Renewable and Sustainable Energy Reviews*, *55*, 1210–1224. https://doi.org/10.1016/j.rser.2015.03.099

Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence*, *112*(November 2021), 104860. https://doi.org/10.1016/j.engappai.2022.104860

Gohel, P., Singh, P., & Mohanty, M. (2021). *Explainable AI: current status and future directions*. 1–16. http://arxiv.org/abs/2107.07045

Haifaa Nasser Hussein, Noor hashim Hamed, O. Z. jasim. (2023). A python-based model for the selection of suitable locations for installing large-scale photovoltaic system in western Iraq by using GIS-machine learning techniques. *The Fourth International Conference on Civil and Environmental Engineering Technologies, June 2020*. https://doi.org/10.1063/5.0140830 ?

Halder, B., Banik, P., Almohamad, H., Al Dughairi, A. A., Al-Mutiry, M., Al Shahrani, H. F., & Abdo, H. G. (2022). Land Suitability Investigation for Solar Power Plant Using GIS, AHP and Multi-Criteria Decision Approach: A Case of Megacity Kolkata, West Bengal, India. *Sustainability (Switzerland)*, *14*(18). https://doi.org/10.3390/su141811276

Halder, B., Banik, P., Almohamad, H., Dughairii, A. A. Al, Al-Mutiry, M., Shahrani, H. F. Al, & Abdo, H. G. (2021). Land suitability analysis for solar farms exploitation using the GIS and Analytic Hierarchy Process (AHP) – a case study of Morocco. *Polityka Energetyczna – Energy Policy Journal*, *24*(2), 79–96. https://doi.org/10.33223/epj/133474

Hernandez, J. M. (2018). *Suitable location for renewable energy using GIS and a financial approach . Study case Valle del Cauca – Colombia*. 1–10.

Hou, Y., Wang, Q., & Tan, T. (2023). Regional suitability assessment for straw-based power generation: A machine learning approach. *Energy Strategy Reviews*, *49*(August), 101173. https://doi.org/10.1016/j.esr.2023.101173

Husein, H. N., Hamed, N. H., & Jasim, O. Z. (2023). A Python-Based Model for the Selection of Suitable Locations for Installing Large-Scale Photovoltaic System in Western Iraq by Using GIS-Machine Learning Techniques. *AIP Conference Proceedings*, *2775*(1), 050009. https://doi.org/10.1063/5.0140830

Islam, M. K., Hassan, N. M. S., Rasul, M. G., Emami, K., & Chowdhury, A. A. (2022). Assessment of solar and wind energy potential in Far North Queensland, Australia. *Energy Reports*, *8*, 557–564. https://doi.org/10.1016/j.egyr.2022.10.134

Itza Alejandra, Mário Sílvio Rochinha de Andrade Caetano, Bañón, Pla, F., & Hugo Alexandre Gomes da Costa. (2020). *Landcover and Crop Type Classification*. University of Nova, Lisbon, Portugal.

Khandakar, A., Chowdhury, M. E. H., Kazi, M. K., Benhmed, K., Touati, F., Al-Hitmi, M., & Gonzales, A. S. P. (2019). Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar. *Energies*, *12*(14). https://doi.org/10.3390/en12142782

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *2017-Decem*(December), 4766–4775.

Maher, A. M., Abdullah, S. H., Ramli, M. F., & Ash'aari, Zu. H. (2015a). A Review of Land Suitability Analysis for Urban Growth by using the GIS-Based Analytic Hierarchy Process. *Asian Journal of Applied Sciences*, *December*, 2321–0893.

Maher, A. M., Abdullah, S. H., Ramli, M. F., & Ash'aari, Zu. H. (2015b). Review of Land Suitability Analysis for Urban Growth by using the GIS-Based Analytic Hierarchy Process. *Asian Journal of Applied Sciences*, *December*, 2321–0893.

Munkhbat, U., & Choi, Y. (2021). Gis-based site suitability analysis for solar power systems in Mongolia. *Applied Sciences (Switzerland)*, *11*(9). https://doi.org/10.3390/app11093748

Nzelibe, I. U., Ojediran, D. D., & Moses, M. (2022). Geospatial Assessment and Mapping of Suitable Sites for a Utility-scale Solar PV Farm in Akure South, Ondo State, Nigeria. *Geomatics and Environmental Engineering*, *16*(4), 79–101. https://doi.org/10.7494/geom.2022.16.4.79

Oyewola, O. M., Ismail, O. S., Olasinde, M. O., & Ajide, O. O. (2022). Mapping of solar energy potential in Fiji using an artificial neural network approach. *Heliyon*, *8*(7), e09961. https://doi.org/10.1016/j.heliyon.2022.e09961

Pranav R, Shashank T K, G. H. S. (n.d.). *Machine Learning for Predicting Solar Power Output*. https://time.com/3723592/inside-the-worlds-largest-solar-power-plant/

Rangel-Martinez, D., Nigam, K. D. P., & Ricardez-Sandoval, L. A. (2021). Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. *Chemical Engineering Research and Design*, *174*, 414–441. https://doi.org/10.1016/j.cherd.2021.08.013

REN21. (2022). Renewables 2022 Global Status. In *Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector*. https://www.ren21.net/gsr-2022/

Rida, A., Hicham, A., & Ilias Kacimi. (2017). Photovoltaic Site Suitability Analysis sing Analytical Hierarchy Process and Sensitivity Analysis Methods with GIS and Remote Sensing in Southern Morocco. *International Renewable and Sustainable Energy Conference (IRSEC)*, *2018*. https://doi.org/10.1109/IRSEC.2017.8477325

Rida Azmi, Hicham Amar, I. K. (2017). Photovoltaic Site Suitability Analysis sing Analytical Hierarchy Process and Sensitivity Analysis Methods with GIS and Remote Sensing in Southern Morocco. *IEEE Geoscience and Remote Sensing Letters*, *2017*. https://doi.org/978-1-5386-2847-8/17

Sachit, M. S., Shafri, H. Z. M., Abdullah, A. F., Rafie, A. S. M., & Gibril, M. B. A. (2022a). Global Spatial Suitability Mapping of Wind and Solar Systems Using an Explainable AI-Based Approach. *ISPRS International Journal of Geo-Information*, *11*(8). https://doi.org/10.3390/ijgi11080422

Sachit, M. S., Shafri, H. Z. M., Abdullah, A. F., Rafie, A. S. M., & Gibril, M. B. A. (2022b). Global Spatial Suitability Mapping of Wind and Solar Systems Using an Explainable AI-Based Approach. *ISPRS International Journal of Geo-Information*, *11*(8). https://doi.org/10.3390/ijgi11080422

Saha, S., & Mondal, P. (2022). Estimation of the effectiveness of multi-criteria decision analysis and machine learning approaches for agricultural land capability in Gangarampur Subdivision, Eastern India. Artificial Intelligence in Geosciences, 3(December), 179–191. https://doi.org/10.1016/j.aiig.2022.12.003

Sahin, G., Isik, G., & van Sark, W. G. J. H. M. (2023). Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Reports*, *10*(September), 2837–2849. https://doi.org/10.1016/j.egyr.2023.09.097

Sánchez-Lozano, J. M., García-Cascales, M. S., & Lamata, M. T. (2015). Evaluation of suitable locations for the installation of solar thermoelectric power plants. *Computers and Industrial Engineering*, *87*(2015), 343–355. https://doi.org/10.1016/j.cie.2015.05.028

Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, *7*(April), 100230. https://doi.org/10.1016/j.dajour.2023.100230

Sarkodie, W. O., Ofosu, E. A., & Ampimah, B. C. (2022). Decision optimization techniques for evaluating renewable energy resources for power generation in Ghana: MCDM approach. *Energy Reports*, *8*, 13504–13513. https://doi.org/10.1016/j.egyr.2022.10.120

Shahab, A., & Singh, M. P. (2019). Comparative Analysis of Different Machine Learning Algorithms in Classification of Suitability of Renewable Energy Resource. *International Conference on Communication and Signal Processing 2019*, *2019*. https://doi.org/10.1109/ICCSP.2019.8697969

Sharifzadeh, M., Sikinioti-Lock, A., & Shah, N. (2019). Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews*, *108*(March), 513–538. https://doi.org/10.1016/j.rser.2019.03.040

Shrestha, S., & Vanneschi, L. (2018). *IMPROVED FULLY CONVOLUTIONAL NETWORK WITH CONDITIONAL RANDOM FIELD FOR BUILDING EXTRACTION Sanjeevan Shrestha IMPROVED FULLY CONVOLUTIONAL NETWORK WITH CONDITIONAL RANDOM FIELD FOR*. New university of Lisbon, Portugal.

Singh, A. S. and M. P. (2019). Comparative Analysis of Different Machine Learning Algorithms in Classification of Suitability of Renewable Energy Resource. *International Conference on Communication and Signal Processing*, 257–263. https://doi.org/10.1109/BDICN55575.2022.00057

Sirefelt Rickard. (2004). *Semi-automatic road extraction from aerial images* [CHALMERS UNIVERSITY OF TECHNOLOGY]. https://doi.org/10.1117/12.508365

Spyridonidou, S., & Vagiona, D. G. (2023). A systematic review of site-selection procedures of PV and CSP technologies. *Energy Reports*, *9*, 2947–2979. https://doi.org/10.1016/j.egyr.2023.01.132

Sudha, S., Edwin, D. F. X., & Nivetha, M. (2023). Integrated Machine Learning Algorithms and MCDM Techniques in Optimal Ranking of Battery Electric Vehicles. E3S Web of Conferences, 405. https://doi.org/10.1051/e3sconf/202340502005

Sun, Y., Zhu, D., Li, Y., Wang, R., & Ma, R. (2023a). Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. *Energy Conversion and Management*, *289*(November 2022), 117198. https://doi.org/10.1016/j.enconman.2023.117198

Sun, Y., Zhu, D., Li, Y., Wang, R., & Ma, R. (2023b). Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. *Energy Conversion and Management*, *289*(May 2023), 117198. https://doi.org/10.1016/j.enconman.2023.117198

Sun, Y., Zhu, D., Li, Y., Wang, R., & Ma, R. (2023c). Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. *Energy Conversion and Management*, *289*(May), 117198. https://doi.org/10.1016/j.enconman.2023.117198

Suprova, N. T., Zidan, R., & Rashid, A. R. M. H. (2020). Optimal Site Selection for Solar Farms Using GIS and AHP: A Literature Review. *Proceedings of the International Conference on Industrial & Mechanical Engineering and Operations Management*, 1–13.

Taoufik, M., Laghlimi, M., & Fekri, A. (2021). Land suitability analysis for solar farms exploitation using the GIS and Analytic Hierarchy Process (AHP) – A case study of Morocco. *Polityka Energetyczna*, *24*(2), 79–96. https://doi.org/10.33223/epj/133474

Teruel-Solano, J., Sánchez-Lozano, J. M., Soto-Elvira, P. L., & Socorro García-Cascales, M. (2013). Geographical Information Systems (GIS) and Multi-Criteria Decision Making (MCDM) methods for the evaluation of solar farms locations: Case study in south-eastern Spain. *Renewable and Sustainable Energy Reviews*, *24*, 544–556. https://doi.org/10.1016/j.rser.2013.03.019

Tisza, K. (2014). *Gis-Based Suitability Modeling and Multi-Criteria Decision Analysis for Utility Scale Solar Plants in Four States in the Southeast US* (Issue May) [Clemson University]. https://tigerprints.clemson.edu/all_theses/2005/?utm_source=tigerprints.clemson.edu%2Fall_theses%2F2005&utm_medium=PDF&utm_campaign=PDFCoverPages

U.S. Renewable Energy Factsheet. (2023). *U.S. PHOTOVOLTAIC INSTALLATIONS, 2011-2022*. University Og Michigan. https://css.umich.edu/publications/factsheets/energy/us-renewable-energy-factsheet

Vajda, S., Kuhn, H. W., & Tucker, A. W. (1951). Contribution to the Theory of Games. In *Journal of the Royal Statistical Society. Series A (General)* (Vol. 114, Issue 4, p. 577). https://doi.org/10.2307/2981092

Villacreses, G., Martínez-Gómez, J., Jijón, D., & Cordovez, M. (2022). Geolocation of photovoltaic farms using Geographic Information Systems (GIS) with Multiple-criteria decision-making (MCDM) methods: Case of the Ecuadorian energy regulation. *Energy Reports*, *8*, 3526–3548. https://doi.org/10.1016/j.egyr.2022.02.152

Wang, C. N., Dang, T. T., Nguyen, N. A. T., & Wang, J. W. (2022). A combined Data Envelopment Analysis (DEA) and Grey Based Multiple Criteria Decision Making (G-MCDM) for solar PV power plants site selection: A case study in Vietnam. *Energy Reports*, *8*, 1124–1142. https://doi.org/10.1016/j.egyr.2021.12.045

Watson, J. J. W., & Hudson, M. D. (2015). Regional Scale wind farm and solar farm suitability assessment using GIS-assisted multi-criteria evaluation. *Landscape and Urban Planning*, *138*, 20–31. https://doi.org/10.1016/j.landurbplan.2015.02.001

Yousefi, H., Hafeznia, H., & Yousefi-Sahzabi, A. (2018). Spatial site selection for solar power plants using a gis-based boolean-fuzzy logic model: A case study of Markazi Province, Iran. *Energies*, *11*(7). https://doi.org/10.3390/en11071648

Annexes A

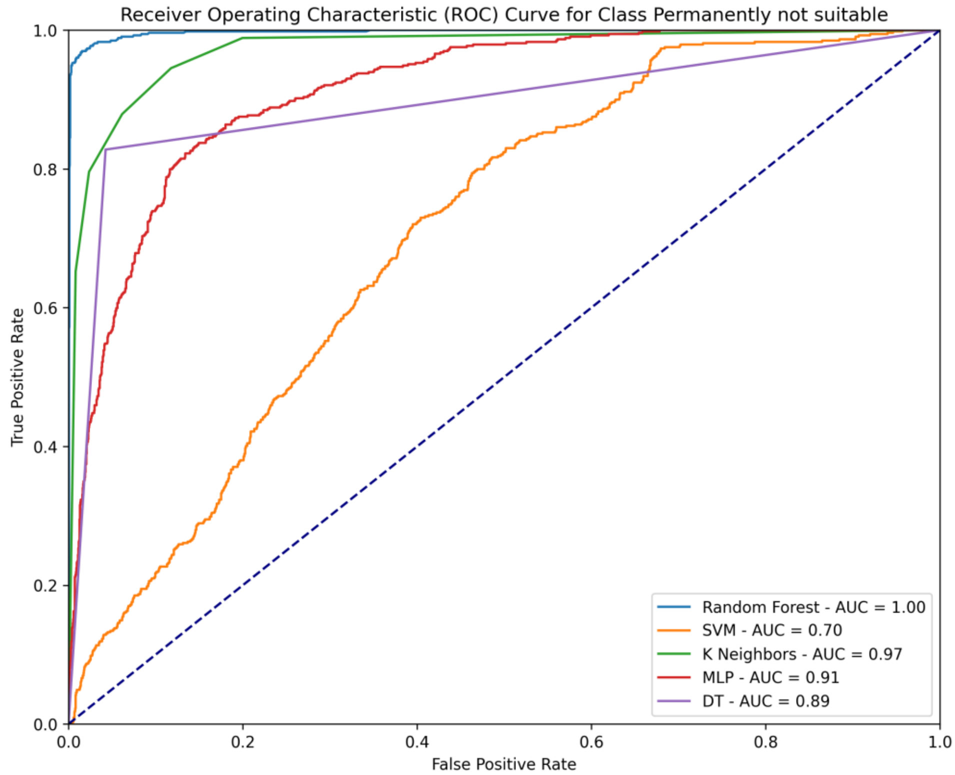

Figure A.2: ROC curve for permanently not suitable class

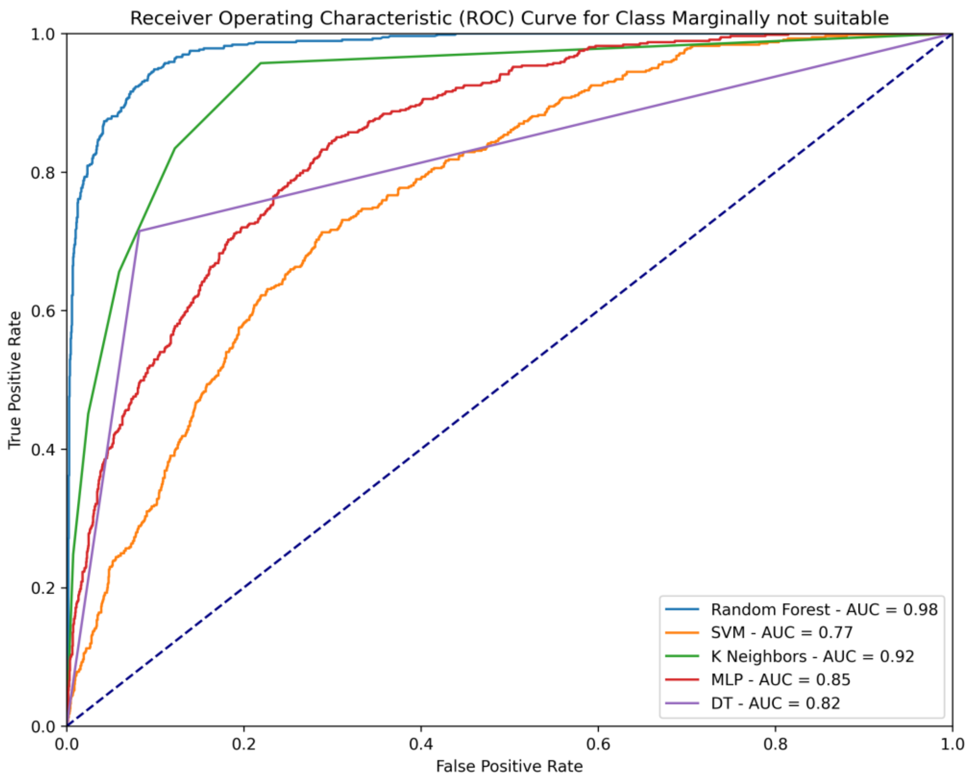

Figure A.1: ROC curve for marginally not suitable class

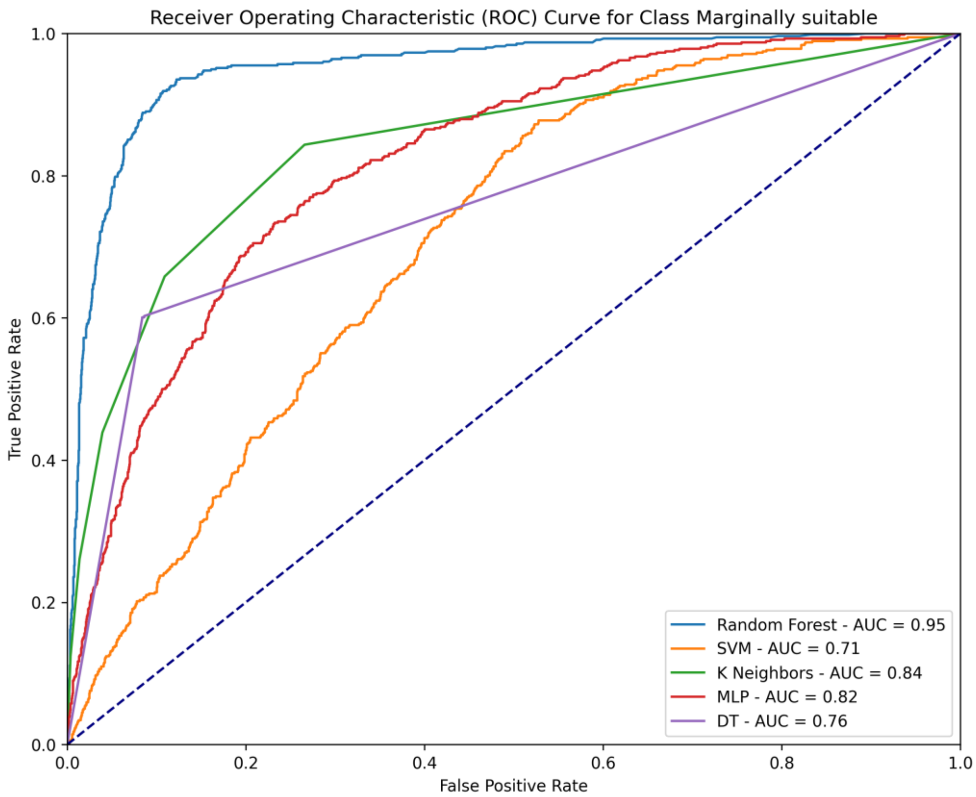

Figure A.4: ROC curve for marginally suitable class

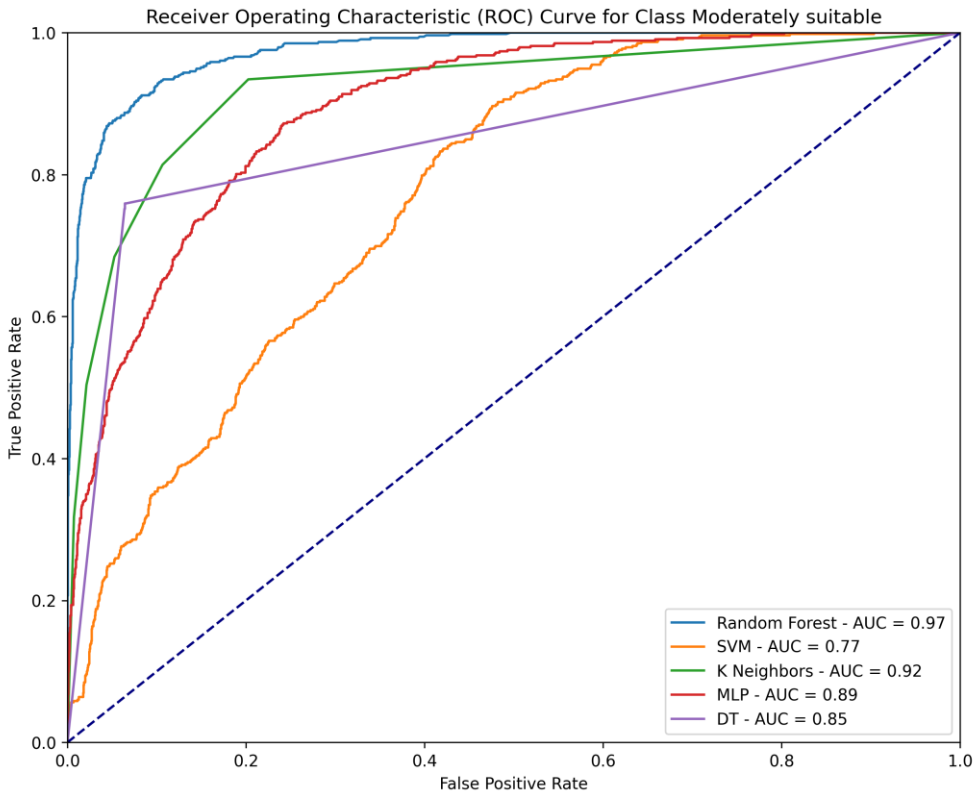

Figure A.3: ROC curve for moderately suitable class

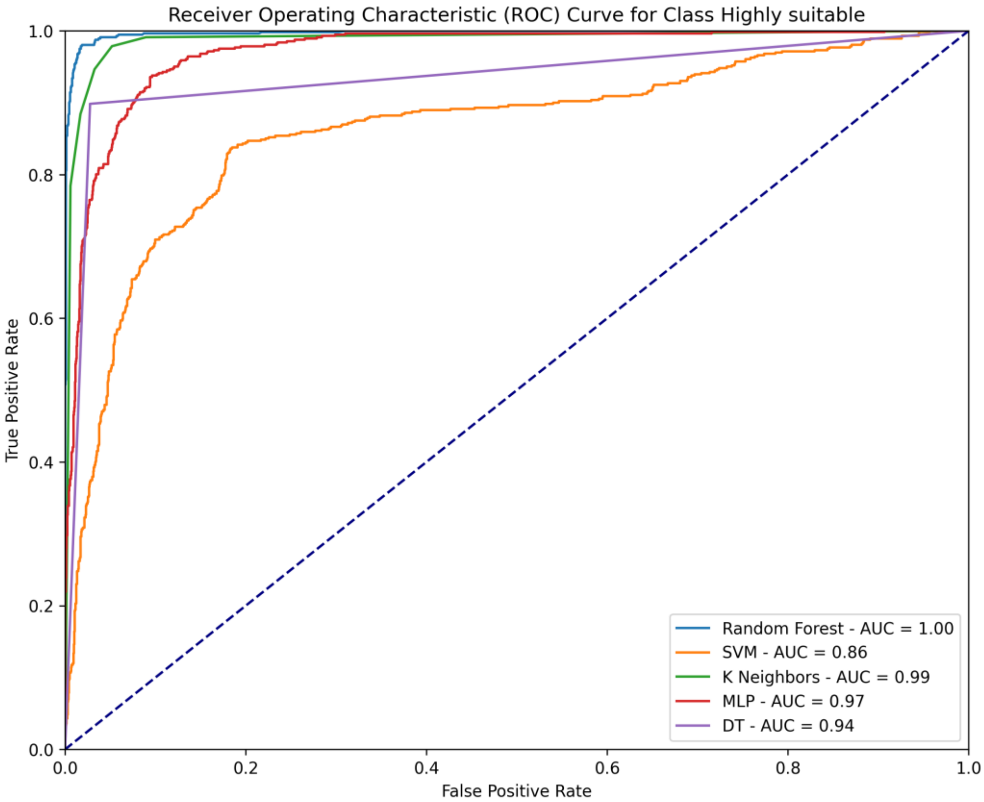

Figure A.5: ROC curve for highly suitable class

Annexes B

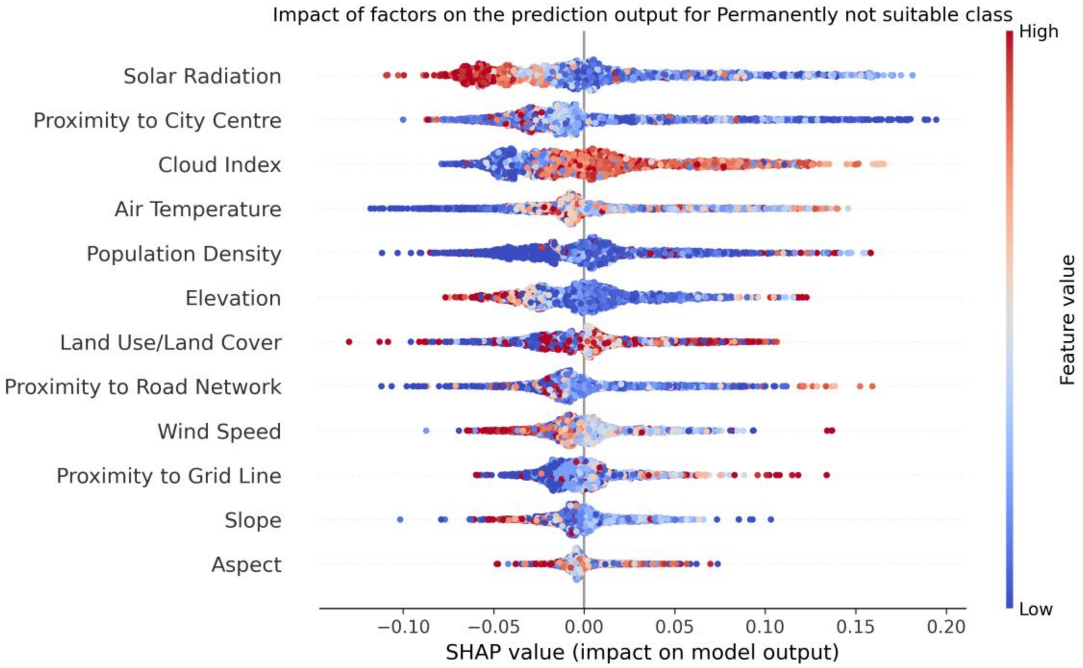

Figure B.1: Significance of conditioning factors for the predictions of permanently not suitable class

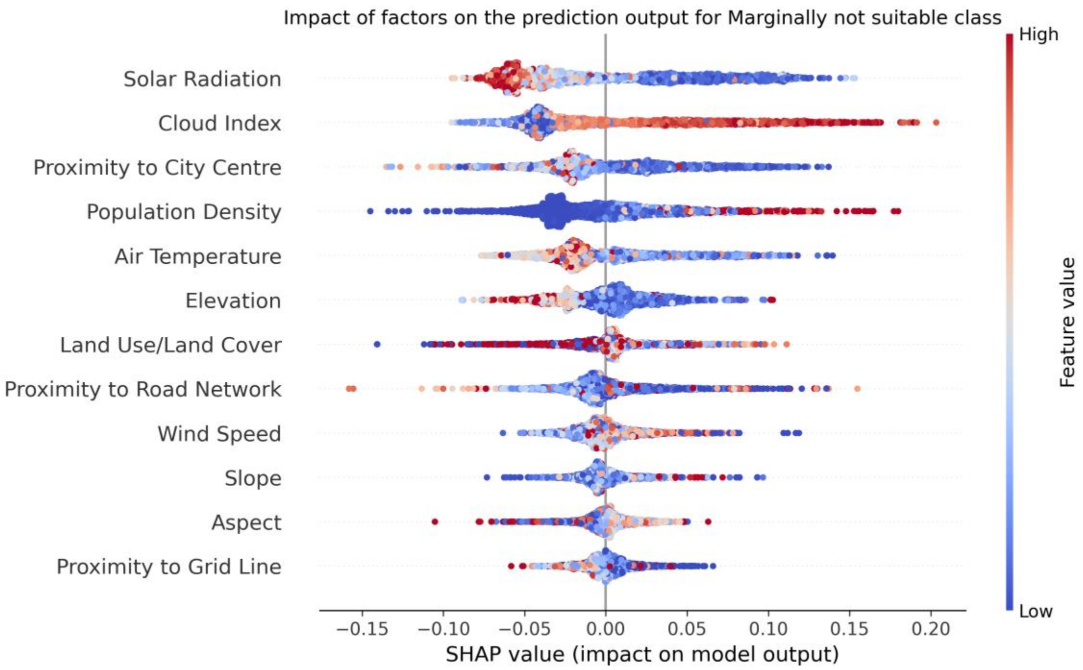

Figure B.2: Significance of conditioning factors for the predictions of marginally not suitable class

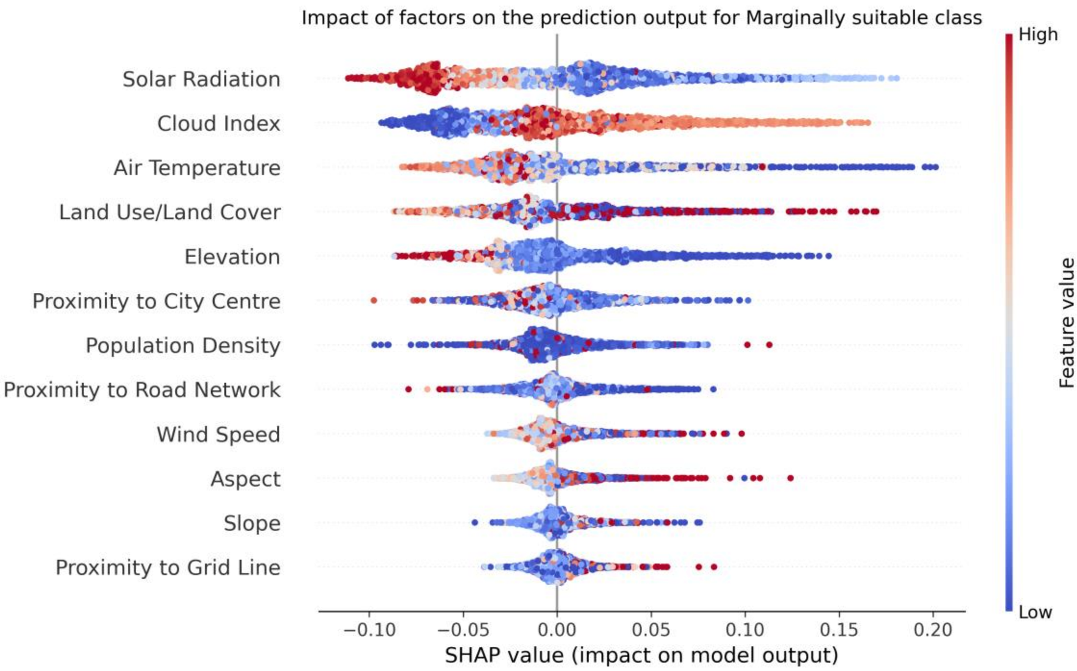

Figure B.3: Significance of conditioning factors for the predictions of marginally suitable class

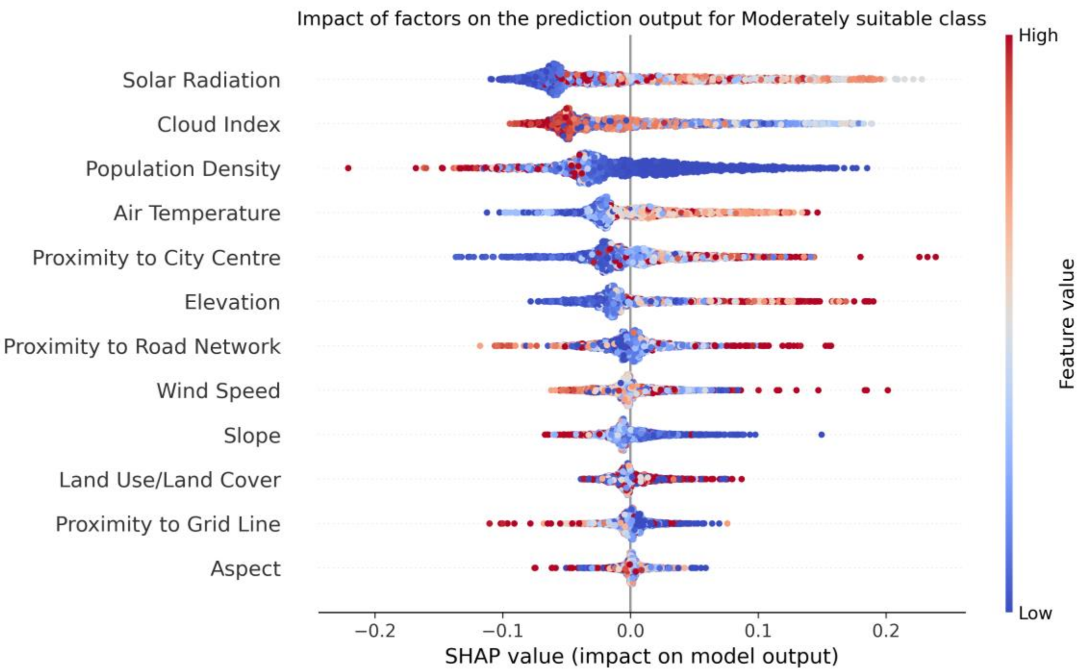

Figure B. 4: Significance of conditioning factors for the predictions of highly suitable class

Masters
Program
in **Geospatial**
Technologies

Supported by:

Education and Culture

ERASMUS MUNDUS