Elucidating the Relationship Between Climate Change and Poverty using Graph Neural Networks, Ensemble Models, and Remote Sensing Data

Parinthapat Pengpun*
Bangkok Christian International School
Bangkok, Thailand
parinzee@protonmail.com

Alessandro Salatiello
University of Tübingen
Tübingen, Germany
alessandro.salatiello@uni-tuebingen.de

Abstract

Climate and poverty are intrinsically related: regions with extreme temperatures, large temperature variability, and recurring extreme weather events tend to be ranked among the poorest and most vulnerable to climate change. Nevertheless, there currently is no established method to directly estimate the impact of specific climate variables on poverty and to identify geographical regions at high risk of being negatively affected by climate change. In this work, we propose a new approach based on Graph Neural Networks (GNNs) to estimate the effect of climate and remote sensing variables on poverty indicators measuring Education, Health, Living Standards, and Income. Furthermore, we use the trained models and perturbation analyses to identify the geographical regions most vulnerable to the potential variations in climate variables.

1 Introduction

Climate change, a global exigency, affects not just the natural world but has profound socio-economic implications as well [1]. From altering agricultural outputs [2] to affecting public health [3, 4], the tentacles of climate change reach far and wide. One such significant socio-economic metric affected by these changes is poverty [5, 6]. Understanding this relationship and its extent can provide a foundation for creating informed policies that holistically address both the environment and socio-economic health [7]. By identifying areas with higher risk of falling into poverty from climate change we hope that we will help raise awareness and allow for preventative measures to be implemented. [8]

While numerous studies have delved into poverty analysis using various indicators and surveys, few have leveraged extensive open climate data exclusively without any localized datasets[9, 10, 11, 12]. This paucity becomes even more pronounced when one considers the application of methodologies like Graph Neural Networks (GNNs) for such an analysis. Furthermore, even though many works have explored the usage of machine learning models to predict poverty [13, 14, 15, 16] few have attempted to clarify the relationships between the features used and the predicted poverty indexes.

Our research fills this gap by systematically analyzing the relationship between open climate data and poverty indexes — leveraging state-of-the-art models, including GNNs and ensemble models, and explainability methods, such as SHapley Additive exPlanations (SHAP) [17].

*Also affiliated with DeepInvest & DeepVision

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

2 Related Work

In the quest to understand and predict poverty, traditional methodologies have predominantly leaned on socio-economic surveys, census data, and labor-intensive ground truthing processes [9, 10, 11, 12]. While providing a detailed picture, these methods are time-consuming, resource-intensive, and can sometimes lag behind rapidly changing realities, particularly in developing nations where socio-economic conditions can swiftly fluctuate [18, 19, 20].

With the advent of data-driven methodologies, recent years have seen a surge in the application of machine learning models to predict poverty. Such endeavors often utilize private or localized datasets as features ranging from household surveys to mobile phone data [14, 15, 16]. However, while powerful, many of these datasets are region-specific, limiting their scalability across borders.

Graph Neural Networks (GNNs) have gained prominence in various domains, from molecular chemistry [21] and social network analysis [22] to recommendation systems [23] and physical systems [24]. Their adaptability to dynamic regions of interest, size invariance, and scalability make them particularly suited for complex tasks [25, 26, 27, 28] where the data have an inherent graph structure [23, 29].

Explainable AI (XAI) has emerged as a critical area of research, particularly in applications where understanding the decision-making process is as important as the decision itself [30, 31, 32, 33]. Methods like LIME (Local Interpretable Model-agnostic Explanations) [34] and SHAP (SHapley Additive exPlanations) [17] have been employed to interpret complex models, providing insights into feature importance and decision rationale [35, 36, 37, 38]. In the context of poverty prediction and climate factors, explainability can offer valuable insights into which variables are most influential, thereby guiding policy interventions more effectively [7, 39, 40, 41].

One paradigm within XAI that has shown promise is perturbation-based methods. These methods investigate the properties of machine learning models by altering the input in various ways—such as occluding part of an image or replacing a word in a sentence—and observing the resulting changes in the model's output [42]. Perturbation-based methods have been applied across various data types, including images [43, 44], videos [45, 46], and natural language [47]. These methods offer several advantages: they provide an intuitive way to explore black-box models, allow for dynamic analysis, and are generally applicable to any models [42, 48].

t

3 Methodology

3.1 Data

The primary dataset used in this study consists of open-access climate data obtained from various sources, including the National Oceanic and Atmospheric Administration (NOAA), European Centre for Medium-Range Weather Forecasts (ECMWF), Copernicus Sentinel-2, MODIS, and more. The dataset includes variables such as temperature, precipitation, humidity, and wind speed from the year 2022. Additionally, poverty indexes for Thailand at a sub-district level were obtained from TP-MAP (Thai People Map & Analytics Platform) [49]. Out of 5 poverty indexes, we focus our experiments on 4 of them: Education, Living Standards, Health, and Income. All data were preprocessed to handle missing values, and to normalize the scales. Feature selection was performed using correlation analysis to remove multi-collinear variables — often found to be detrimental to model performance [50, 51, 52]. A complete list and explanation of the features used, TP-MAP Dataset, and data preprocessing steps can be found in the Appendices A.1, A.2, and A.3, respectively.

3.2 Models

We leveraged a two-stage pipeline with two tasks: classification and regression. The first-stage model classifies whether or not each sub-district is affected by poverty, while the second-stage model performs regression to predict the degree to which poverty has affected the sub-district. We train all classification models using Stratified KFold cross-validation, and train all regression models using KFold cross-validation. Classification model performance is evaluated using Accuracy and F1. Regression model performance is evaluated using Mean Absolute Error (MAE), Mean Squared

Error (MSE), and Root Mean Square Error (RMSE). Due to the nature of our data, we considered Ensemble Models (often state-of-the-art on tabular data) and GNN models (often state-of-the-art on graph data).

To establish a baseline for comparison, we incorporated a "dummy model" in our study. This model uses simple heuristics: for classification, it predicts the most common class (mode) from the training dataset, and for regression, it predicts the average (mean) of the training labels. This model's predictions are uniform, ignoring input features, providing a basic benchmark against our more sophisticated models.

3.2.1 Ensemble Models

We utilize ensemble models due to their effectiveness with tabular data, crucial for analyzing our climate and poverty indices. These models handle diverse data types well, capture complex relationships, and offer straightforward interpretability through feature importance scores. Their inclusion provides a robust baseline for comparison, enriching our understanding of the key drivers of poverty in the context of climate change and validating the effectiveness of our modeling approach.

To build strong ensemble models, we considered popular ensemble models including LightGBM [53], XGBoost [54], CatBoost [55], and Random Forest [56], and other traditional baselines including Support Vector Machines (SVM) [57], and K-nearest neighbors (KNN) [58]. We then tuned the hyper-parameters of the top 5 models and combined them in a stacking and voting ensemble.

3.2.2 Graph Neural Networks

The selection of Graph Neural Networks (GNNs) for our study is driven by their strong alignment with the challenges of analyzing climate change impacts on poverty. GNNs are particularly adept at spatial data representation, crucial for our geographically-oriented dataset [23, 29]. They excel in relational learning, enabling them to capture complex interdependencies between regions [25, 26, 27, 28], a key aspect in understanding how climate factors influence poverty. Their scalability is essential given the extensive data involved in our research. This combination of features makes GNNs an ideal choice for effectively analyzing and predicting the multifaceted impacts of climate change on poverty.

We tested various GNN architectures such as Graph Convolution Network (GCN), GraphSAGE, Graph Attention Network (GAT), ChebNet, and Graph Isomorphism Network (GIN). Nodes in the graph represent geographical regions in Thailand with attributes extracted from the climate dataset, while edges are unweighted. We tuned each architecture's hyper-parameters extensively for both tasks: aggregation functions, number of hidden layers, activation functions, dropout rates, initialization function, and others. For a comprehensive list of tuned hyper-parameters please see Appendix A.4.

3.3 Model Explanation

To study the relationship between climate variables and poverty indices for tabular models, we apply the SHapley Additive exPlanations (SHAP) [17] method to the best-performing models. This method allows us to estimate the magnitude and direction of the effect of each feature on each considered target poverty index.

For GNNs, we utilize GNNExplainer [59]. This tool effecively identifies critical node features pertaining to the GNN's prediction. Its application is vital in our context, as it not only demystifies the decision-making process of GNNs but also ensures that our model's predictions are based on meaningful and relevant data patterns. GNNExplainer aids in validating the model's performance and informs further model refinements, aligning our study with the demands for transparent and accountable AI, especially in areas with profound socio-economic and environmental implications.

To identify the sub-districts at the highest risk of being affected by climate change, we employ a perturbation-based approach. Specifically, we leverage best-performing models to measure the changes in poverty metrics in response to changes in individual climate factors. For example, we may increase the temperature values for a particular sub-district and assess how this perturbation affects the predicted poverty.

Table 1: Performance Metrics for Different Models on Dataset Targets

	Education		Health		Income		Living Std.	
Model	CLS	REG	CLS	REG	CLS	REG	CLS	REG
Best GNN	**.62/.80**	**.19/.24**	**.63/.76**	**.22/.25**	**.86**/.59	**.23/.27**	**.63/.76**	**.22/.25**
Best Ensemble	.52/.77	.21/.25	.54/.68	.24/.28	.36/.76	.23/.27	.52/.64	.23/.27
Dummy	.00/.73	.25/.29	.00/.70	.25/.29	.00/**.85**	.25/.29	.00/.67	.25/.29

[1] CLS: Classification scores are reported as (F1/Accuracy).
[2] REG: Regression scores are reported as (MAE/RMSE).

4 Results

Classification Models The performance of various models on different dataset targets is summarized in Table 1. For classification tasks, the F1 score and accuracy were used as the evaluation metrics. Our Graph Neural Network (GNN) model outperformed the best ensemble model across all dataset targets. Specifically, in the education domain, our GNN model achieved an F1 score of **.62** and an accuracy of **.80**, compared to .52 and .77 for the best ensemble model. Similar trends were observed in the health, income, and living standards domains, reinforcing the efficacy of our GNN model for classification tasks.

Regression Models For regression tasks, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used as the evaluation metrics. Once again, our GNN model demonstrated superior performance. In the education domain, the GNN model achieved an MAE of **.19** and an RMSE of **.24**, which are better than the best tabular model's .21 and .25, respectively. This trend of outperformance by our GNN model was consistent across all dataset targets.

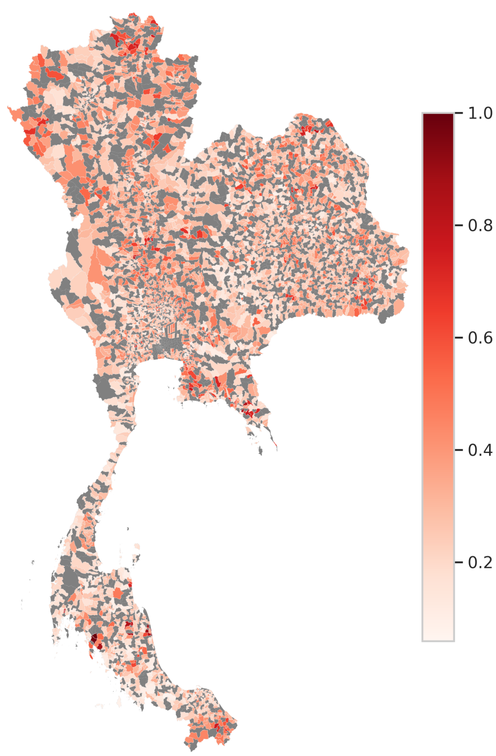

Figure 1: Sensitivity Map Obtained from Perturbation Analysis done on the best GNN on the Education Poverty Index. The higher the score the more sensitive that sub-district is to climate change. The gray sub-districts are ones where we don't have the training/testing data for.

5 Discussion

5.1 Model Explanation

To further understand the inner workings of our models, we employed SHAP and GNNExplainer for the tabular and GNN model respectively, to interpret their predictions. The plots revealed that certain features had a more significant impact on the model's decision-making process than others. For instance, NO2 levels, Aerosol Optical Thickness, and evaporation are most influential features for predicting poverty rates. Furthermore, it is evident that temperatures, pollution (temperature of O3), and crops also play a big role. The plots and more details are available in Appendix A.5.

5.1.1 Identifying Vulnerable Sub-districts through Perturbation Analysis

By introducing small perturbations to the data and observing the changes in the model's predictions, we were able to pinpoint rows that were particularly sensitive to changes. Specifically, we identified that the most sensitive areas are among the North and Northwest parts of Thailand. Identifying such vulnerable instances is crucial for targeted interventions and resource allocation.

This could be due to the fact that in the mountainous regions of North and Northwest Thailand, educational access is severely hindered by geographical isolation, difficult terrain, and infrastructural deficiencies, including sporadic internet and electricity provision [60]. The COVID-19 pandemic has intensified these challenges, disrupting education systems and disproportionately affecting marginalized communities, with girls and young women at particular risk of educational discontinuity due to increased domestic responsibilities during the crisis [60].

6 Conclusion

This study presents a comprehensive analysis of the intricate relationship between climate change and poverty in Thailand, leveraging advanced machine learning techniques such as Graph Neural Networks (GNNs) and ensemble models, along with perturbation analsysis and other XAI methods. The core achievement of our research is the enhanced understanding it offers of the socio-economic impacts of climate change, providing valuable insights for stakeholders in policy-making and resource allocation such as governments and NGOs. By utilizing open climate data and GNN models, our approach surpasses traditional methods, offering a more dynamic, accurate, and granular analysis of poverty-stricken areas. This is complemented by the robustness and comprehensive nature of the predictions made by ensemble models, which together with GNNs, provide a well-rounded analysis tool.

In conclusion, our research marks a significant step forward in using technology to understand and address the challenges of climate change on socio-economic conditions. It opens up new avenues for research and practical application, aiming to create more informed, effective, and equitable interventions. We hope that our study will not only contribute to academic discourse but also translate into real-world impact, aiding in the development of resilient communities better equipped to handle the ramifications of climate change.

Acknowledgments and Disclosure of Funding

The authors would like to extend their sincere appreciation to the Artificial Intelligence Association of Thailand (AIAT) and the Super AI Engineer Season 3 program for graciously providing the detailed TPMAP data originally sourced from their hackathon. This data was instrumental in our research and analysis. We also acknowledge the creators of the Thai People Map and Analytics Platform (TPMAP)—developed by the Office of National Economic and Social Development Board (NESDB), National Electronics and Computer Technology Center (NECTEC), and National Science and Technology Development Agency (NSTDA) under the Ministry of Science and Technology. Their efforts in developing such a comprehensive tool have significantly contributed to precision poverty alleviation and quality of life improvements for Thai citizens. Their dedication to creating a data-driven solution for identifying and addressing the unique challenges faced by impoverished communities across Thailand has been invaluable to our work.

References

[1] Shubhi Patel, Anwesha Dey, Shani Kumar Singh, Rakesh Singh, and H.P. Singh. *Socio-Economic Impacts of Climate Change*, chapter 12, pages 237–267. John Wiley Sons, Ltd, 2021.

[2] David B. Lobell and Sharon M. Gourdji. The influence of climate change on global crop productivity. *Plant Physiology*, 160(4):1686–1697, 2012.

[3] I. Orimoloye, S. Mazinyo, Ahmed Mukalazi Kalumba, O. Y. Ekundayo, and W. Nel. Implications of climate variability and change on urban and human health: A review. *Cities*, 2019.

[4] C. Cooney. Climate change and infectious disease: Is the future here? *Environmental Health Perspectives*, 119:a394 – a397, 2011.

[5] S. Hallegatte, M. Fay, and E. Barbier. Poverty and climate change: introduction. *Environment and Development Economics*, 23:217 – 233, 2018.

[6] R. Leichenko and Julie A. Silva. Climate change and poverty: vulnerability, impacts, and alleviation strategies. *Wiley Interdisciplinary Reviews: Climate Change*, 5, 2014.

[7] Tristan Ballard, Gopal Erinjippurath, Matthew W Cooper, and Chris Lowrie. Widespread increases in future wildfire risk to global forest carbon offset projects revealed by explainable ai. In *ICLR 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023.

[8] Alec Crawford, Simon Mason, Andrew Kruczkiewicz, and Pietro Ceccato. Accessing and using climate data and information in fragile, data-poor states. `https://www.iisd.org/publications/report/accessing-and-using-climate-data-and-information-fragile-data-poor-states`, 2015.

[9] Kannika Thampanishvong, Jerasorn Santisirisomboon, Boonwara Sumano, Charika Channuntapipat, Kanyaphak Ngaosri, Patree Witoonchart, Nitchakarn Kaewbuadee, and Lokchan Pinkaew. Impact assessment of climate change and environmental degradation on children in thailand. Technical report, Oct 2022.

[10] World Bank Group. Thailand rural income diagnostic: Challenges and opportunities for rural farmers. Technical report, World Bank, Bangkok, October 2022.

[11] OECD. Oecd economic surveys: Thailand 2020: Economic assessment. Technical report, Paris, 2020.

[12] National Statistics Office Thailand (NSO). The 2022 household socio-economic survey. Technical report, 2023.

[13] Nattapong Puttanapong, Arturo Martinez, Joseph Albert Nino Bulan, Mildred Addawe, Ron Lester Durante, and Marymell Martillan. Predicting poverty using geospatial data in thailand. *ISPRS International Journal of Geo-Information*, 11(5), 2022.

[14] Jessica E. Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqbal, Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Tatem, and Linus Bengtsson. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127):20160690, 2017.

[15] N. Pokhriyal and D. Jacques. Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 114:E9783 – E9792, 2017.

[16] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature Communications*, 11(1):2583, 5 2020.

[17] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[18] Faizuddin Ahmed, Cheku Dorji, Shinya Takamatsu, and Nobuo Yoshida. *Hybrid Survey to Improve the Reliability of Poverty Statistics in a Cost-Effective Manner*. The World Bank, 2014.

[19] Jinyao Lin, Shuyi Luo, and Yiqin Huang. Poverty estimation at the county level by combining luojia1-01 nighttime light data and points of interest. *Geocarto International*, 37:3590 – 3606, 2021.

[20] J. Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350:1073 – 1076, 2015.

[21] Oliver Wieder, Stefan Kohlbacher, Mélaine Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Seidel, and Thierry Langer. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 37:1–12, 2020.

[22] Qiaoyu Tan, Ninghao Liu, and Xia Hu. Deep representation learning for social network analysis. *Frontiers in Big Data*, 2, 2019.

[23] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation, 2019.

[24] Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, and Peter Battaglia. Graph networks as learnable physics engines for inference and control, 2018.

[25] Luana Ruiz, Fernando Gama, and Alejandro Ribeiro. Graph neural networks: Architectures, stability, and transferability. *Proceedings of the IEEE*, 109:660–682, 2020.

[26] A. Sun, P. Jiang, M. Mudunuru, and Xingyuan Chen. Explore spatio-temporal learning of large sample hydrology using graph neural networks. *Water Resources Research*, 57, 2021.

[27] Fernando Gama, E. Isufi, G. Leus, and Alejandro Ribeiro. Graphs, convolutions, and neural networks: From graph filters to graph neural networks. *IEEE Signal Processing Magazine*, 37:128–138, 2020.

[28] Lilin Cheng, Haixiang Zang, Zhi nong Wei, Tao Ding, and Guo qiang Sun. Solar power prediction based on satellite measurements – a graphical learning method for tracking cloud motion. *IEEE Transactions on Power Systems*, 37:2335–2345, 2022.

[29] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: A survey, 2022.

[30] Waddah Saeed and Christian Omlin. Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263:110273, 2023.

[31] J. Amann, A. Blasimme, E. Vayena, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(310), 2020.

[32] D. Shin. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *Int. J. Hum. Comput. Stud.*, 146:102551, 2021.

[33] Sheikh Rabiul Islam, William Eberle, Sheikh Khaled Ghafoor, and Mohiuddin Ahmed. Explainable artificial intelligence approaches: A survey, 2021.

[34] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier, 2016.

[35] Edoardo Mosca, Ferenc Szigeti, Stella Tragianni, Daniel Gallagher, and Georg Groh. SHAP-based explanation methods: A review for NLP interpretability. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4593–4603, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.

[36] Michele Cafagna, Lina Maria Rojas-Barahona, Kees van Deemter, and Albert Gatt. Interpreting vision and language generative models with semantic visual priors. *ArXiv*, abs/2304.14986, 2023.

[37] Letitia Parcalabescu and Anette Frank. Mm-shap: A performance-agnostic metric for measuring multimodal contributions in vision and language models tasks, 2023.

[38] Carlos Miguel Chiesa-Estomba, Manuel Graña, Bader Aldughayfiq, Farzeen Ashfaq, Noor Zaman Jhanjhi, and Mamoona Humayun. Explainable ai for retinoblastoma diagnosis: Interpreting deep learning models with lime and shap. *Diagnostics*, 13, 2023.

[39] Alona Zharova, Annika Boer, Julia Knoblauch, Kai Ingo Schewina, and Jana Vihs. Explainable multi-agent recommendation system for energy-efficient decision support in smart homes. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022.

[40] Maike Sonnewald, Redouane Lguensat, Aparna Radhakrishnan, Zoubero Sayibou, Venkatramani Balaji, and Andrew Wittenberg. Revealing the impact of global warming on climate modes using transparent machine learning and a suite of climate models. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.

[41] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning, 2019.

[42] Maksims Ivanovs, Roberts Kadikis, and Kaspars Ozols. Perturbation-based methods for explaining deep neural networks: A survey. *Pattern Recognition Letters*, 150:228–234, 2021.

[43] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer, 2014.

[44] Ruth C Fong and Andrea Vedaldi. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE international conference on computer vision*, pages 3429–3437, 2017.

[45] Zhenqiang Li, Weimin Wang, Zuoyue Li, Yifei Huang, and Yoichi Sato. Towards visually explaining video understanding networks with perturbation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1120–1129, 2021.

[46] Joonatan Manttari, Sofia Broomé, John Folkesson, and Hedvig Kjellstrom. Interpreting video features: A comparison of 3d convolutional networks and convolutional lstm networks. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[47] Shusen Liu, Zhimin Li, Tao Li, Vivek Srikumar, Valerio Pascucci, and Peer-Timo Bremer. Nlize: A perturbation-driven visual interrogation tool for analyzing and interpreting natural language inference models. *IEEE transactions on visualization and computer graphics*, 25(1):651–660, 2018.

[48] Ruth Fong, Mandela Patrick, and Andrea Vedaldi. Understanding deep networks via extremal perturbations and smooth masks. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2950–2958, 2019.

[49] Office of National Economic, Social Development Board, National Electronics, and Computer Technology Center. Thai People Map and Analytics Platform, 2023. Available at `https://www.tpmap.in.th/`.

[50] Jong Hae Kim. Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72:558 – 569, 2019.

[51] Christopher Winship and Bruce Western. Multicollinearity and model misspecification. *Sociological Science*, 3:627–649, 2016.

[52] D. Montgomery and D. J. Friedman. Prediction using regression models with multicollinear predictor variables. *Iie Transactions*, 25:73–85, 1993.

[53] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[54] Tianqi Chen and Carlos Guestrin. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016.

[55] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. Catboost: gradient boosting with categorical features support, 2018.

[56] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[57] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[58] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[59] Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks, 2019.

[60] United Nations in Thailand. The more you learn, the more you earn: education and poverty alleviation in thailand. United Nations in Thailand, 2023. Available online: `https://thailand.un.org/` (accessed on December 1, 2023).

A Appendices

A.1 Data Features from Google Earth Engine

Table 2: List of Features Extracted from Google Earth Engine (Aggregated over 2022)

Dataset	Features
Sentinel-2 Multispectral Instrument	B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12, AOT, WVP
Sentinel-5P Aerosol	absorbing_aerosol_index
Sentinel-5P Carbon Monoxide	CO_column_number_density, H2O_column_number_density
Sentinel-5P Formaldehyde	tropospheric_HCHO_column_number_density, tropospheric_HCHO_column_number_density_amf, HCHO_slant_column_number_density
Sentinel-5P Nitrogen Dioxide	NO2_column_number_density, tropospheric_NO2_column_number_density, stratospheric_NO2_column_number_density, NO2_slant_column_number_density, tropopause_pressure
Sentinel-5P Ozone	O3_column_number_density, O3_effective_temperature
Sentinel-5P Sulfur Dioxide	SO2_column_number_density, SO2_column_number_density_amf, SO2_slant_column_number_density
Sentinel-5P Methane	CH4_column_volume_mixing_ratio_dry_air
MODIS Leaf Area Index	Fpar, Lai
MODIS Vegetation Indexes	NDVI, EVI
MODIS Land Surface Temperature	LST_Day_1km, LST_Night_1km
MODIS Evapotranspiration	ET, LE, PET, PLE
JAXA GSMaP Hourly Precipitation	hourlyPrecipRateGC
ECMWF ERA5 Climate	dewpoint_temperature_2m, temperature_2m, skin_temperature, soil_temperature_level_1, soil_temperature_level_2, soil_temperature_level_3, soil_temperature_level_4, volumetric_soil_water_layer_1, volumetric_soil_water_layer_2, volumetric_soil_water_layer_3, volumetric_soil_water_layer_4, forecast_albedo, evaporation_from_bare_soil_sum, evaporation_from_open_water_surfaces_excluding_oceans_sum, evaporation_from_the_top_of_canopy_sum, evaporation_from_vegetation_transpiration_sum, runoff_sum, surface_pressure, total_precipitation_sum, leaf_area_index_high_vegetation, leaf_area_index_low_vegetation, surface_latent_heat_flux_sum, surface_net_solar_radiation_sum, surface_net_thermal_radiation_sum, surface_sensible_heat_flux_sum, surface_solar_radiation_downwards_sum, surface_thermal_radiation_downwards_sum
Google Dynamic World	water, trees, grass, flooded_vegetation, crops, shrub_and_scrub, built, bare
FIRMS Fire Radiative Power	T21
NOAA VIIRS Nighttime Lights	avg_rad

A.2 Details on TP-MAP Dataset

Thai People Map and Analytics Platform (TPMAP) [49] is a data analytics tool developed by the Office of National Economic and Social Development Board (NESDB) and National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Ministry of Science and Technologies. The platform aims to provide precision poverty alleviation and improve the quality of life for Thai citizens by integrating data from various government agencies.

A.2.1 Multidimensional Poverty Index Indicators

TPMAP employs the Multidimensional Poverty Index (MPI) developed by Oxford Poverty & Human Development Initiative and United Nation Development Programme (UNDP). The MPI uses five dimensions to identify poverty: healthcare, education, income, living standard, and access to public services. Below are the indicators used for each dimension:

Table 3: Multidimensional Poverty Index Indicators in TPMAP

Dimension	Indicators
Health Care	Newborn weight above 2.5 KG Food consumption meets hygienic standards, Proper use of medicines, Regular exercise (3 times a week, 30 minutes each)
Living Standard	Safe housing conditions, Access to drinking water (5 liters/person/day), Access to clean water for daily use (45 liters/person/day), House kept tidy and hygienic
Education	Proper care for children age 3-5, Mandatory nine-years education for children age 6-14, Continuation to higher education levels, Literacy and basic math skills for age 15-59
Income	Proper jobs and income for age 15-59, Average income per year of individual household members
Access to Public Services	Proper care for elders, Proper care for the disabled

A.3 Data Preprocessing and Splitting

This appendix provides a detailed explanation of the data preprocessing and splitting steps carried out in this study.

A.3.1 Preprocessing

1. **Aggregation:** We aggregate our features by taking the mean of each over the temporal axis. In other words, we take the average of each feature over the whole year.
2. **Removal of Multi-colinear Features:** Features with a correlation coefficient greater than 0.8 were identified. Among the correlated pairs, the feature deemed less important according to the best model at that time was removed.
3. **Iterative Feature Imputation:** Missing features in the dataset were filled using an ensemble of the best-performing tabular models in an iterative manner.
4. **Feature Standardization:** All features were transformed to have a mean of 0 and a standard deviation of 1, ensuring that no particular feature dominates the model.
5. **Target Variable Transformation:** For regression tasks, the target variable underwent two transformations. First, it was transformed to follow a uniform distribution. Then, its values were scaled to lie within a 0-1 range.

A.3.2 Data Splitting For Classification and Regression

1. **Two-Stage Model:** The study employed a two-stage modeling approach, consisting of a classification model followed by a regression model.

2. **Classification:** The first stage involved training a classification model to predict whether a sub-district is affected by poverty or not.
3. **Regression:** The second stage involved training a regression model, but only on sub-districts with non-zero poverty indices.

A.4 Hyperparameters in GNN Architectures

This appendix provides a comprehensive list of hyperparameters tuned in the Graph Neural Network (GNN) architectures used in this study.

Table 4: Hyperparameters for GNN Architectures

Hyperparameter	Description	Possible Values
Activation Function	Type of activation function	ReLU, LeakyReLU, ELU
Aggregation Function	Function to aggregate neighbor information	add, mean, max, min, softmax, powermean
Number of Layers	Total number of GNN layers	[3, 12]
Dropout Probability	Probability of dropping out a unit	[0.1, 0.7]
Hidden Size	Number of hidden units	[64, 512]
Attention Heads	Number of attention heads (GAT only)	[1, 10]
Chebyshev Polynomial Degree	Degree of Chebyshev polynomial (Cheb only)	[2, 5]
Weight Initialization	Method for initializing weights	Xavier, He
Skip Connections	Whether to use skip connections	True, False
Normalization Type	Type of normalization layer	Off, BatchNorm, LayerNorm

A.5 Feature Importance Plots from SHAP and GNNExplainer

The combined feature importance and SHAP value analyses from the Graph Neural Network (GNN) classifier and tabular model reveal crucial environmental and socio-economic predictors that influence poverty in Thailand. Environmental factors such as *evaporation_from_vegetation_transpiration* and *forecast_albedo* emerge as highly influential in the GNN classifier, suggesting a strong link between natural resource indicators and educational outcomes. The SHAP value analysis complements this finding, highlighting the varied impacts of *stratospheric_NO2_column_number* and *AOT*, indicative of complex interactions with levels of poverty. Other features, notably *evaporation_from_open_water* and *LST_Day_1km*, display predominantly negative impacts on poverty predictions, hinting at their associations with more affluent conditions.

These analytical insights unveil the intricate interplay between ecological variables and poverty, emphasizing the necessity for a nuanced understanding in policy formulation and resource allocation. The interpretability afforded by SHAP values provides a transparent picture of the contributing factors, which is crucial for making informed decisions in socio-economic planning. Such detailed analyses offer a clear pathway for stakeholders to strategically target interventions and for researchers to delve deeper into the causal dynamics of poverty, ensuring that efforts are both effective and efficiently directed.

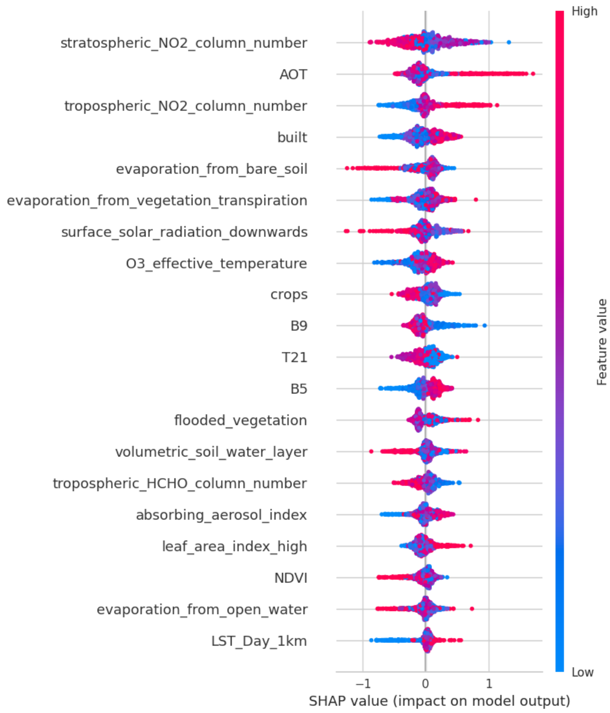

Figure 2: SHAP Interpretation of Education Poverty Index Classification Model

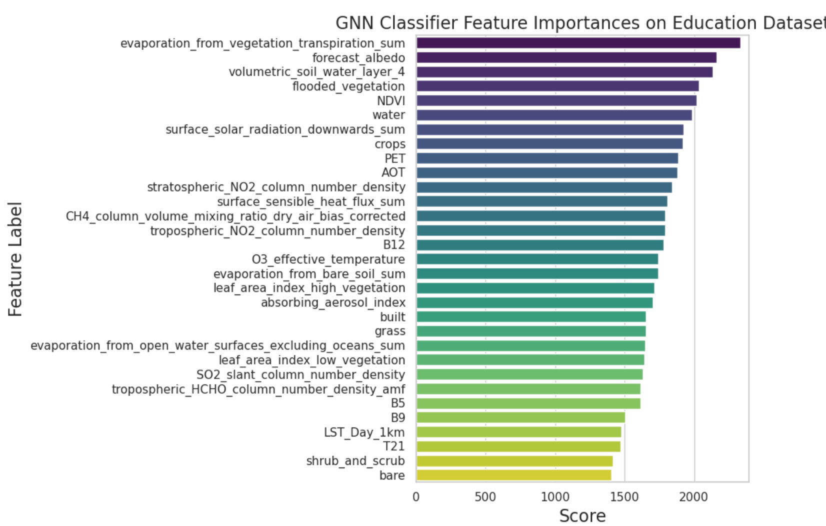

Figure 3: GNNExplainer Interpretation of Education Poverty Index Classification Model