**I-GUIDE MODEL CARD**

The I-GUIDE Model Card is an easy-to-use tool that will allow you to create documentation for each model that you create or use in a project.

Using this tool will help facilitate transparency and reproducibility about your project. It will also help you comply with relevant policies of journals, funding agencies, and universities.

The Model Card applies to:

* Pre-existing models acquired from other sources, e.g., produced by other researchers;
* Models you and your collaborators produced yourselves;
* Models you and your collaborators produced by integrating two or more other models (e.g., coupling).

**Model Card Attribution**

This Model Card template is an adapted version of the I-GUIDE Data Card template, which itself is based on Google’s *Data Cards Playbook* (https://pair-code.github.io/datacardsplaybook/).  
It has been restructured to address key considerations for geospatial model transparency, performance evaluation, and ethical deployment, in alignment with the I-GUIDE research lifecycle.

This work is shared under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).  
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

A logo with text on it

AI-generated content may be incorrect.

* + 1. **BASIC INFORMATION**

|  |  |
| --- | --- |
| Model Card ID Number | *FD-01* |
| Model Name | *FloodDepth-MultiBranchLearning* |
| Model Version | *0.01* |
| Persistent Identifier | *N/A* |
| Outputs Supported | *N/A* |
| Model Card Author | *Jikun Liu, Texas A&M University, jikun@tamu.edu* |

* + 1. **MODEL OVERVIEW**

|  |  |  |
| --- | --- | --- |
| Model Type | *(Select all that apply)*  ☐ AI model: Feature Routing Deep Learning Model  ☐ Statistical model: *(Specify type)*  ☐ Other: *(Specify)* | |
| Purposes | *(Select all that apply)*  ☐ Classification  ☐ Decision support  ☐ Forecasting  ☐ Regression  ☐ Simulation  ☐ Spatial analysis  ☐ Other: *(Specify)* | |
| Domains of Application | *(Select all that apply)*  ☐ Climate science  ☐ Economics  ☐ Environmental impact modeling  ☐ Geospatial analysis  ☐ Hydrology  ☐ Population modeling  ☐ Other social systems modeling: *(Specify)*  ☐ Other: *(Specify)* | |
| Model Authors and Developers | *Jikun Liu, Texas A&M University,* [*jikun@tamu.edu*](mailto:jikun@tamu.edu) | |
| Source and Acquisition Method | ☐ Acquired (from external source)  ☐ Developed internally  ☐ Integrated from multiple models (e.g., coupled) | |
| User Licensing | ☐ Open source: *(Specify license type)*  ☐ Proprietary: *(Specify owner)*  ☐ Other restrictions on use: *(Specify restrictions)* | |
| Storage Location | ☐ Repository: *https://github.com/VarianJou/Flood-height-prediction*  ☐ Project-specific storage: *(Describe location)* |
| Access Control Policies | ☐ Open  ☐ Embargoed: *(Describe release timeline)*  ☐ Restricted: *(Describe access criteria)* |
| Use Case | *This model was designed to allow multi-branch learning for flood depth prediction at given geolocation so that the original dataset could be broken down into smaller characteristic subsets and thus learn different sets of parameters for better abstraction, achieving better overall performance.* | |

* + 1. **MODEL INPUTS AND TRAINING DATA**

|  |  |
| --- | --- |
| Model Inputs | *Unless otherwise specified, all following data are point-aggregated:*  *Weather data: maxwind\_kph\_day, wind\_kph, wind\_degree, pressure\_mb, precip\_mm, humidity, dewpoint\_c, uv, gust\_kph, feelslike\_c, windchill\_c, heatindex\_c, chance\_of\_rain, chance\_of\_snow, vis\_km on point of prediction.*  *Timestamps: year, month, day, hour*  *DEM: dem\_min,dem\_max,dem\_mean,dem\_iqr*  *Impervious surface coverage: total\_area\_km2, pct\_area\_1, pct\_area\_2, area\_km\_1, area\_km\_2, cai\_1, cai\_2*  *Land Cover: pct\_area\_forest, area\_km\_forest, cai\_forest, pct\_area\_shrubland, area\_km\_shrubland, cai\_shrubland, pct\_area\_herbaceous, area\_km\_herbaceous, cai\_herbaceous, pct\_area\_planted, area\_km\_planted, cai\_planted, pct\_area\_wetlands, area\_km\_wetlands, cai\_wetlands*  *Sentinel-1 RS: VV\_Min, VV\_Max, VV\_Mean, VV\_IQR, VV\_SD, VH\_Min, VH\_Max, VH\_Mean, VH\_IQR, VH\_SD, VH\_VV\_Ratio*  *Other: soil\_moisture, longitude, latitude* |
| Input Data Types | *(Select all that apply)*  ☐ Raster  ☐ Tabular  ☐ Time Series  ☐ Vector  ☐ Other: *(Specify)* |
| Training Data Used | *Weather data mentioned above is collected from API* [*https://www.weatherapi.com/*](https://www.weatherapi.com/)*.*  *The prediction target – HighWaterMark with timestamps – is collected from USGS, check the links below for reference:* [*https://www.usgs.gov/publications/identifying-and-preserving-high-water-mark-data*](https://www.usgs.gov/publications/identifying-and-preserving-high-water-mark-data) *and* [*an example query*](https://maps.waterdata.usgs.gov/mapper/nwisquery.html?URL=https://waterdata.usgs.gov/usa/nwis/current?search_station_nm_match_type=anywhere&state_cd=tx&index_pmcode_STATION_NM=1&index_pmcode_DATETIME=2&format=sitefile_output&sitefile_output_format=xml&column_name=agency_cd&column_name=site_no&column_name=station_nm&sort_key_2=site_no&html_table_group_key=NONE&rdb_compression=file&list_of_search_criteria=state_cd&column_name=site_tp_cd&column_name=dec_lat_va&column_name=dec_long_va&column_name=agency_use_cd)  *DEM and Land Cover data are collected from USGS*  *Sentinel-1 is collected from Copernicus Data Space Ecosystem (CDSE)* |
| Training Dataset Representativeness | *The dataset is known to be highly sparse across the US, even given a relatively large timeframe. More target data are available on rural environments instead of urban ones.* |

* + 1. **MODEL STRUCTURES**

|  |  |
| --- | --- |
| Feature Selection | *All aforementioned datasets in the Model Inputs section were used.* |
| Hyperparameters and Tuning | *Grid search should be considered the most important parameter tuning to allocate the best fitting branch-structure combination; an experimentalist approach was employed to check which grouping of features combined with which selections of parameters were the most effective.* |
| Software and Dependencies | *The model is trained with widely used Python libraries such as pytorch and sklearn* |

* + 1. **MODEL PERFORMANCE AND EVALUATION**

|  |  |
| --- | --- |
| Validation Approach | ☐ Cross-validation  ☐ Holdout set  ☐ Time series split  ☐ Other: *(Specify)* |
| Evaluation Results | *0.253 MSE on designated regression task for flood depth* |
| Testing or Validation Data Used | The validation data is a partition of the original dataset. |
| *(If model is integrated from multiple other models)*  Contribution of Constituent Models | *N/A* |

* + 1. **MODEL ADAPTATION AND CUSTOMIZATION (for acquired or integrated models only)**

|  |  |
| --- | --- |
| Source Models | *N/A* |
| Availability of Source Model Code | *N/A* |
| Modifications | *N/A* |
| Training Data Adjustments | *N/A* |

* + 1. **MODEL DEPLOYMENT AND USAGE**

|  |  |
| --- | --- |
| Computational Requirements | *No known requirements.* |
| Geospatial Considerations | *Trained with US data. Migrating this model could cause issues.* |

* + 1. **TRANSPARENCY, EXPLAINABILITY, AND INTERPRETABILITY**

|  |  |
| --- | --- |
| Model Transparency | ☐ Fully transparent (rule-based, interpretable ML)  ☐ Partially transparent (some explainability features)  ☐ Black box (deep learning, complex ML models) |
| Explainability Features | ☐ Feature importance analysis  ☐ LIME  ☐ Sensitivity analysis  ☐ SHAP values  ☐ Other: *Feature routing by itself should be considered an explainability feature because its performance would only be better when similar features were combined and fed into known structures, thereby leaving room for characteristics generalization* |
| Interpretability Challenges | *How to interpret grouping and structural tuning results remains hard* |
| Communication of Model Limitations | *No uncertainty measurement is provided. The user would be able to explicitly see which features are routed through which branch.* |

* + 1. **OTHER ETHICAL CONSIDERATIONS**

|  |  |
| --- | --- |
| Ethical Risks (Other Than Transparency, Explainability and Interpretability) | *(Select all that apply)*  ☐ Bias in training data: *Rural focus*  ☐ Intentional misuse risks: *(Specify)*  ☐ Privacy risks and surveillance: *(Specify)*  ☐ Security risks: *(Specify)*  ☐ Stigmatization of individuals or communities: *(Specify)*  ☐ Other: *(Specify)* |
| Measures Taken to Address Ethical Risks | *Stratified Train-valid Split* |
| Suitable Uses | *This model would be suitable for US flood depth prediction using described datasets where it is known to be a flood.* |
| Unsuitable Uses | *This model has yet to run a specificity test, and thus it is unknown how well it would perform in non-flood periods. It is not yet ready for use cases such as alarm system that require minimal false-positive.*  *This model does not distinguish between rural and urban setting specifically, and its performance could diverge significantly in either setting. Use with caution for urban environments.* |