A Hybrid Rf-Symbolic Regression Approach for Accurate Solar Irradiance Prediction in Mountain Regions

Aleksandr GEVORGIAN[1]*, Andrea GASPARELLA[2], Giovanni PERNIGOTTO[3]

[1]Free University of Bozen-Bolzano, Faculty of Engineering,
Bolzano, Italy
Tel: +39 0471 017083, Fax: +39 0471 017983, email: aleksandr.gevorgian@natec.unibz.it

[2]Free University of Bozen-Bolzano, Faculty of Engineering,
Bolzano, Italy
Tel: +39 0471 017200, Fax: +39 0471 017009, email: andrea.gasparella@unibz.it

[3]Free University of Bozen-Bolzano, Faculty of Engineering,
Bolzano, Italy
Tel: +39 0471 017632, Fax: +39 0471 017009, email: giovanni.pernigotto@unibz.it

* Corresponding Author

ABSTRACT

This study introduces a novel Hybrid Random Forest-Symbolic Regression model for precise prediction of hourly solar irradiance in mountainous and urban regions. Leveraging geostationary satellite imagery and on-site measurements, the proposed model shows a good performance in accurately predicting solar irradiance, even under challenging conditions such as horizon effects and shading. Notably, it provides interpretable mathematical expressions, enhancing our comprehension of the underlying physical processes. Key findings reveal R^2 values of 0.91 for Global Horizontal Irradiance (*GHI*) and R^2 values ranging from 0.88 to 0.91 regarding different cardinal orientations for Global Tilted Irradiance (*GTI*). Additionally, the model exhibits efficient usage of computational resources, featuring memory utilization between 340.35 MB and 1461.80 MB, execution times spanning 35.41 to 172.89 s, and *CPU* utilization rates from 26.6 % to 37.6 % during mathematical expression generation. This research has the potential to advance the development of environmentally friendly and efficient solar energy systems in terrain-affected regions, aiding in the design of energy-efficient buildings while promoting responsible resource usage.

1. INTRODUCTION

Symbolic regression using Genetic Algorithm (*GA*), a computational process, aims to find simple mathematical expressions that accurately model datasets, offering interpretability advantages over complex models like Random Forest (*RF*) (Breiman, 2001). Unlike traditional regression methods, it seeks to uncover the genuine underlying data generation processes, akin to how physicists derive fundamental expressions for natural phenomena (Schmidt and Lipson, 2009). This involves finding concise and understandable mathematical functions (Alaoui Abdellaoui and Mehrkanoon, 2021).

However, the complexity of the search space in symbolic regression using *GA* (Eiben and Smith, 2003; Banzhaf et al., 1998) grows exponentially when predicting complex phenomena like solar radiation. This results in lengthy and computationally expensive expressions. In response, our method combines *RF* for dataset division with GA-based symbolic regression methods (Stephens, 2016; Banzhaf et al., 1998) to efficiently generate mathematical expressions. This approach simplifies the search space complexity, reducing computational demands and establishing precise relationships between predictors and targets within subsets.

Ultimately, our goal is to predict global horizontal irradiance (*GHI*) and global tilted irradiance (*GTI*) for various cardinal orientations in complex environments, benefiting renewable energy applications like solar *PV* systems, solar water heaters, and building performance analysis.

Our method excels in generating precise mathematical expressions, enabling accurate solar irradiance prediction even in shaded areas. This facilitates data-driven decision-making for urban and mountainous regions, optimizing renewable energy systems and reducing energy costs and emissions. This research contributes to efficient and sustainable solar energy systems in terrain-affected areas and offers insights for energy-efficient building designs.

2. METHODOLOGY

2.1 Case-study location and weather stations

Bolzano, an Italian town nestled in the Alps, is located at approximately 46.500° N latitude and 11.350° E longitude. It spans 30 square kilometers and has a population of around 110,000. The central part of the town sits at an elevation of 262 m above sea level, with surrounding areas ranging from 232 m to over 1600 m.

In this study, we established the first weather station on the rooftop of the E Building at the Free University of Bozen-Bolzano campus in the city center. Equipped with five EKO MS-802 pyranometers, one measures *GHI*, while the remaining four capture *GTI* from diverse cardinal directions with a 90-degree tilt.

The second weather station, located approximately 3 kilometers away within the NOI Techpark in Bolzano, mirrors the configuration of the first station. It features five Delta-T SPN1 Sunshine Pyranometers: alongside one pyranometer positioned horizontally, the other four collect *GTI* with a 90-degree tilt from distinct cardinal directions.

Despite the challenges of mountainous terrain, particularly in the northeast and northwest sectors with elevations reaching approximately 30 degrees above the horizon, the stations have operated since installation in 2017, collecting solar data at a high temporal resolution of 1 minute. Figure 1 presents Bolzano's geography and the weather monitoring equipment used in this study, including topography sourced from topographic-map.com and instruments at the university campus and NOI Techpark.

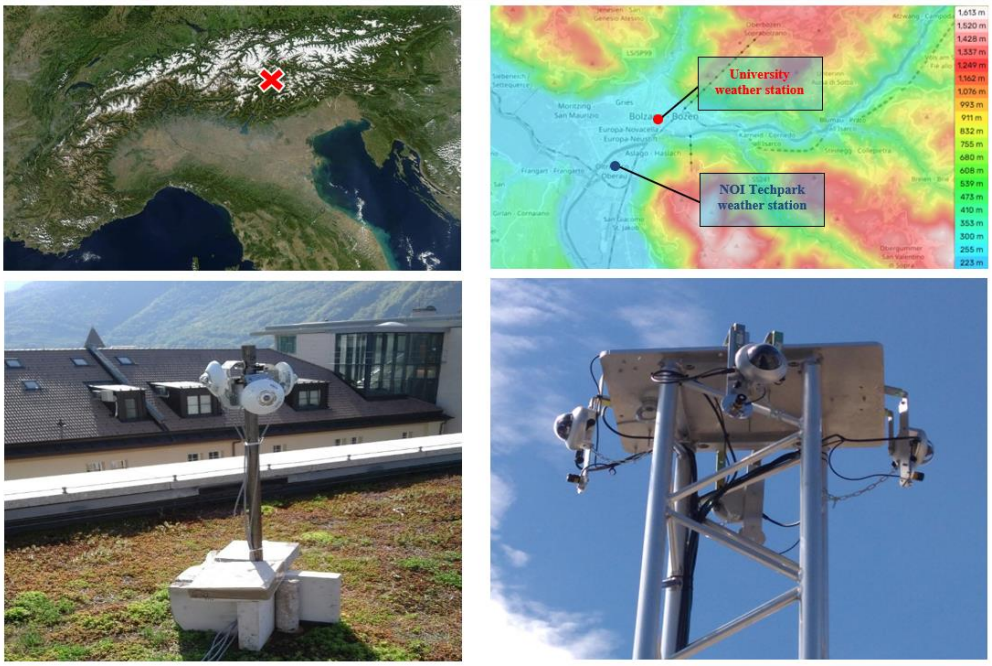

Figure 1: Bolzano overview with topography and weather stations. A comprehensive view of Bolzano featuring Alps topography and terrain data sourced from topographic-map.com. The university campus roof with weather station, including EKO MS-802 pyranometers, and the NOI Techpark weather station with Delta-T SPN1 sunshine pyranometers

2.2 Data collection and preparation

To develop an efficient symbolic regression algorithm for predicting *GHI* and *GTI* across four cardinal orientations, we integrated satellite data (Solcast, 2023) spanning 2019-2022. This included Diffuse Horizontal Irradiance (*DHI*), Direct Normal Irradiance (*DNI*), and Beam Horizontal Irradiance (*BHI*), weather parameters, solar geometry variables, and time-related factors collected hourly for Bolzano. Ground-measured *GHI* and *GTI* was prioritized over satellite ones for precise, location-specific information crucial in complex areas.

The 2019 data pertained to the weather station at the Free University of Bozen-Bolzano's E Building, while the 2022 data were from the NOI Techpark weather station, serving as predictors for our study. Hourly *GHI* and *GTI* measurements from both stations were used as targets.

To ensure the comprehensiveness of our data for this study, we followed a series of preprocessing steps. Initially, we conducted data cleaning to address missing or corrupted data instances (Maharana *et al.*, 2022). This was followed by data normalization (Patro and Sahu, 2015) and the removal of outliers (Nugroho *et al.*, 2021). Additionally, we performed feature selection (Genuer *et al.*, 2010) to identify and retain pertinent predictors in our analysis. For model training, we utilized the 2019 dataset. Following that, we performed model testing using the 2022 dataset (Gonzalez Zelaya, 2019).

2.3 Model training and evaluation

In this study, we aim to efficiently generate mathematical expressions for predicting *GHI* and *GTI* across the four cardinal orientations: North, South, East, and West. To achieve this, we adopt a two-step approach, illustrated in Figure 2.

In the initial step, we employ a *RF* regressor (Breiman, 2001) to convert the dataset into a tree-based structure. This approach presents several advantages. Each decision tree within the *RF* divides the dataset using significant predictor variables, forming distinct tree-based subsets with well-defined predictor-target relationships for *GHI* and *GTI* across the four orientations.

Afterwards, in the second step, we employ a (GA)-based symbolic regression approach, as delineated by Stephens (2016) and Banzhaf et al. (1998), to reveal mathematical expressions within these tree-based subsets. In this context, the *GA* emphasizes smaller, more homogenous data subsets with distinct predictor-target relationships, consequently diminishing the overall complexity of the GA-based symbolic regression model's search space and streamlining the construction of mathematical expressions.

To implement this methodology, we've developed a dedicated Python code. It begins by training a *RF* regressor using the 2019 dataset. The *RF* constructs decision trees for data partitioning and evaluates feature importance. Following this, the code extracts rules from these trees, which represent conditions under which the dataset is partitioned into subsets. The search space of the Symbolic regression (Stephens, 2016) is then guided by these extracted rules to construct precise mathematical expressions representing the relationships between the predictors and target variables within each dataset subset.

In the final phase, we assess the fit quality of each expression within diverse subsets by computing R^2 scores. Top-performing equations for *GHI* and *GTI* with respect to the four cardinal orientations are selected based on superior R^2 scores. These expressions are validated using the entire 2019 training dataset, employing metrics like R^2, *RMSE*, *MAE*, and *MBE*. This comprehensive evaluation process guides the final selection of the most effective expressions.

The coefficient of determination R^2 quantifies the proportion of variance in observed solar radiation explained by the model predictions. A higher R^2 value indicates a better fit of the model to the observed data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{\text{observed},i} - Y_{\text{predicted},i})^2}{\sum_{i=1}^{n}(Y_{\text{observed},i} - \overline{Y_{\text{observed}}})^2} \quad (1)$$

Here, n is the number of data points, $Y_{\text{observed},i}$ represents the observed solar radiation at time i, $Y_{\text{predicted},i}$ represents the predicted solar radiation at time i, and $\overline{Y_{\text{observed}}}$ is the mean of observed solar radiation.

To assess accuracy, we employed *RMSE*, which measures the differences between predicted and observed solar radiation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{\text{observed},i} - Y_{\text{predicted},i})^2} \quad (2)$$

MAE was chosen as a metric to quantify the average absolute difference between predicted and observed solar radiation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_{\text{observed},i} - Y_{\text{predicted},i}\right| \quad (3)$$

MBE was employed to assess the models' bias in predicting solar radiation, indicating whether they systematically overestimate, or underestimate observed values.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{\text{observed},i} - Y_{\text{predicted},i}\right) \quad (4)$$

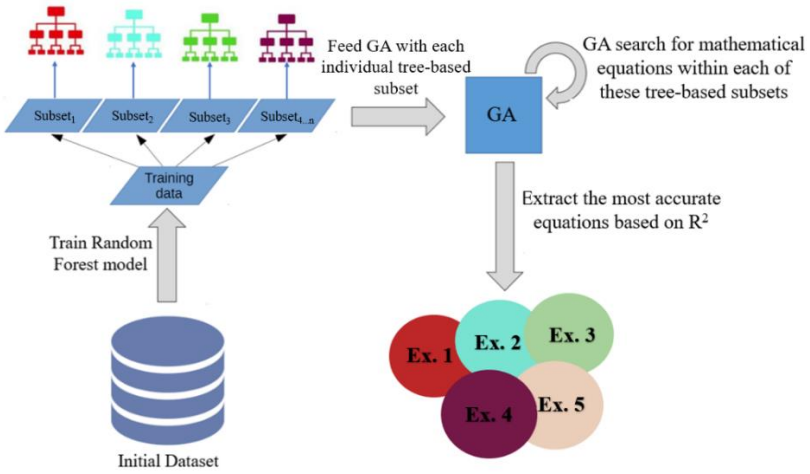

Figure 2: Workflow diagram for enhanced expressions generation

2.4 Performance evaluation

During our performance evaluation, we implemented a calibration procedure to enhance the precision of mathematical expressions used in calculating *GHI* and *GTI*. Our objective was to adapt these expressions to the specific site conditions observed at the NOI Techpark weather station while preserving their structure. This calibration process allows us to use the same equations in various locations, requiring only adjustments to coefficients to suit local conditions.

To achieve this goal, we utilized a custom random search approach, employing Python's *'num_iterations'* function. Within each iteration, we randomly selected coefficients from a uniform distribution ranging from 0.1 to 2.0 for each variable in our mathematical expressions. This aimed to identify the optimal combination of coefficients maximizing expression precision.

To guide our search, we continuously evaluated expression accuracy using a 10% subset of 2022 *GHI* and *GTI* measurements, applying the R^2 metric and selecting coefficients based on the highest R^2 scores. Our comprehensive evaluation compared expression predictions to actual solar radiation data for *GHI* and *GTI* throughout 2022, using metrics like *RMSE*, *MAE*, *MBE*, and R^2 to assess model performance. Calibrating mathematical expressions for specific locations enhances accuracy and adaptability, capturing unique site characteristics and environmental factors for precise modeling. Once calibrated for multiple sites, the model can generalize to similar conditions, aiding renewable energy planning across diverse areas.

2.5 Evaluation of expression generation speed

In our evaluation of the proposed approach, we focused on assessing its computational performance, particularly the time required for generating mathematical expressions and resource utilization. To facilitate this, we developed a custom Python script to streamline experiments and collect performance metrics. It recorded execution time, *CPU* utilization, and peak memory consumption throughout expression generation. The experiment was conducted under consistent, controlled conditions to ensure stable *CPU* utilization and no other running applications. This approach maintained a reliable environment for assessment, minimizing external factors' influence. The script used various libraries to track performance indicators:

- *Memory Usage Tracking*: We utilized the *memory_profiler* library to measure memory usage (Solanki, 2022). This library allowed us to record memory consumption accurately, providing insights into resource utilization.
- *Execution Time Tracking*: Monitoring execution time was achieved using Python's built-in time module (GeeksforGeeks, 2022). It enabled us to capture the start and end times of the symbolic regression process, facilitating the calculation of execution time.
- *CPU Utilization Tracking*: For measuring *CPU* utilization, we employed the *psutil* library (psutil documentation, 2023). This library calculated the percentage of *CPU* utilization during the execution of the *run_code* function, providing insights into *CPU* intensity.

By seamlessly integrating these libraries into our script, we obtained precise measurements of execution time, memory usage, and *CPU* utilization. These metrics delivered valuable insights into how efficient the proposed model utilizes computational resources during the mathematical expression generation process. All experiments were conducted on a laptop computer with specific hardware and software specifications:

- Processor: Intel ® Core™ i5 – 10210U CPU @ 1.60GHz, 2.11GHz
- RAM: 16 GB (15.8 usable)
- Operating System: Windows 10, 64-bit, x64-based processor
- Software Environment: Python 3.11.4 with the *gplearn* library and other relevant libraries for data preprocessing and analysis.

3. RESULTS AND DISCUSSION

3.1 Overview of generated expressions

Our approach has yielded highly promising results in predicting *GHI* and *GTI* across all four cardinal orientations. These outcomes showcase the remarkable capability of our carefully chosen the most effective expressions (expressions 5 to 9) to capture intricate relationships among diverse input and target variables.

Expression 5 (GHI):

$$\mathrm{GHI} = \max\left(\max\left(\max\left(\max\left(\frac{CO}{\min(\Theta_z, RH)}, \frac{1}{CO}\right), \frac{\sqrt{\Theta_{AZ}} - CO \cdot 0.14 \cdot \Theta_{EA} + \max(DP, \Theta_{EA}) + \Theta_{EA} + DP}{\Theta_{EA}}\right), \Theta_{EA}\right), 0.1 \cdot (AM + \Theta_{EA})\right) - AM \quad (5)$$

Expression 5, dedicated to *GHI* estimation, stands as a robust testament to the complexity of solar irradiance modeling. It incorporates a multitude of meteorological variables, including cloud opacity (*CO*), zenith angle (Θ_Z), relative humidity (*RH*), dew point temperature (*DP*), elevation angle (Θ_{EA}), azimuth angle (Θ_{AZ}), and air mass (*AM*). The utilization of nested max and min functions within this expression reflects its adaptability to a wide range of climatic conditions. Notably, it captures the dynamic interplay between these variables, offering a precise estimation of *GHI*.

Expression 6 (GTI_{north}):

$$GTI_{north} = \max\left(\log\left(-(DHI)\right) \cdot \sqrt{GHI}, \min\left(|-(SP)| + (-\sin(GHI)), \sqrt{GHI} + DHI \cdot 0.3\right)\right) \quad (6)$$

Expression 6 tackles the challenging task of *GTI* estimation for north-facing surfaces and incorporates *GHI* computed by the first expression as a critical factor. This expression employs logarithmic, square root, and trigonometric functions, further compounded by the involvement of parameters like *DHI* and Surface Pressure (*SP*). Its complexity mirrors the intricate nature of solar radiation patterns on north-facing surfaces. By employing max and min functions, it navigates the uncertainties posed by varying conditions, providing valuable insights into *GTI*.

Expression 7 (GTI_{west}):

$$GTI_{west} = max\left(max\left(min\left(\sqrt{\left|\left(\Theta_{EA} \cdot min\left((h \cdot h), BHI\right)\right|\right) \cdot BHI}, (h \cdot h)\right), \Theta_{EA} - 0.7\right) + min\left(min\left(min\left(GHI, \left(\frac{(h \cdot h)}{(\Theta_{AZ}}\right) \cdot min\left((h \cdot h), BHI\right)\right), BHI\right), BHI\right), \sqrt{|(\Theta_{EA} \cdot GHI)|}\right) \quad (7)$$

Expression 7 focuses on *GTI* estimation for west-facing surfaces, a domain characterized by its own unique challenges for the considered location of Bolzano. It wields square root functions, combined with parameters including Θ_{EA}, Θ_{AZ}, hours of the day (*h*), *BHI* and computed *GHI*. The nested max and min functions in this expression allow it to navigate the intricacies of solar irradiance dynamics on west-facing surfaces. It captures how factors like time of day, surface orientation, and solar radiation interact, providing valuable insights into *GTI*.

Expression 8 (GTI_{east}):

$$GTI_{east} = min\left(GHI, \left|GHI\sqrt{\frac{|BHI|}{h}}\right|, min(|DHI|, \Theta_z), min\left(\sqrt{GHI \cdot \left(-\sqrt{SP}\right)}, 0.5\sqrt{GHI}\right), \Theta_{AZ}, GHI\right) \quad (8)$$

Expression 8 is designed to estimate *GTI* for east-facing surfaces. This expression incorporates square root, absolute value, and trigonometric functions, coupled with Θ_Z, Θ_{AZ}, *DHI*, computed *GHI* and *BHI*. By employing min functions, it ensures conservative estimates of *GTI*, considering variations in surface orientation and atmospheric conditions.

Expression 9 (GTI_{south}):

$$GTI_{south} = \left|\frac{GHI}{-0.562} \cdot \frac{2AM+\Theta_{EA}}{2}\right| \cdot \sqrt{GHI \cdot max(GHI, max(GHI \cdot DNI \cdot 0.501, DNI)) \cdot 0.501} \quad (9)$$

The final expression is dedicated to *GTI* estimation for south-facing surfaces. Its calculations involve absolute value, multiplication, and square root functions, combined with factors such as *AM*, Θ_{EA}, computed *GHI*, and *DNI*. This expression employs max functions to discern the maximum values among these components, providing insights into the upper limits of *GTI*.

The complexity of the generated expressions aligns with the intricate nature of solar radiation patterns influenced by a multitude of meteorological and geographical factors. These expressions are not just mathematical constructs; they can be powerful tools that offer profound insights into solar irradiance, crucial for optimizing energy systems and harnessing renewable energy efficiently.

3.2 Performance evaluation of the hybrid RF-Symbolic Regression model

In this evaluation, we assess our hybrid RF-Symbolic Regression model's performance in predicting solar irradiance using the entire training dataset from the university weather station and the complete testing dataset from the *NOI* Techpark weather station. Table 1 presents performance metrics covering *GHI* and *GTI* across various cardinal orientations, offering insights into the accuracy and reliability of the model's predictions.

Results from the university weather station dataset show strong correlations between predicted and observed values for *GHI* and *GTI* across different orientations, with R^2 values ranging from 0.88 to 0.91. Low *RMSE* and *MAE* values indicate close alignment between predictions and observations, with *MBE* values close to zero, reflecting balanced predictions without systematic bias.

Performance on the testing dataset from NOI Techpark weather station mirrors that of the training dataset, with high R^2 values (0.91) and consistent *RMSE* and *MAE* values, affirming the expressions' reliability in predicting solar irradiance.

Scatter plots in Figure 3 visually depict the close alignment between our expressions and actual measurements for both *GHI* and *GTI*, further validating their accuracy.

Overall, our evaluation underscores the hybrid model's potential for accurate solar irradiance predictions, offering practical applications in various environments. The unbiased nature of the model enhances its reliability and consistent accuracy, making it a valuable tool for solar energy system engineering.

Table 1: Performance metrics of generated mathematical expressions

University weather station (training dataset)				
Parameter	**R^2**	**RMSE [W m^{-2}]**	**MAE [W m^{-2}]**	**MBE [W m^{-2}]**
GHI	0.91	77.01	44.63	-9.52
GTI Faced North	0.90	16.40	8.34	-0.17
GTI Faced West	0.89	50.27	23.37	-2.02
GTI Faced East	0.89	46.21	17.74	7.71
GTI Faced South	0.88	75.41	35.91	5.04
Noi Techpark weather station (testing dataset)				
Parameter	**R^2**	**RMSE [W m^{-2}]**	**MAE [W m^{-2}]**	**MBE [W m^{-2}]**
GHI	0.91	76.09	45.38	-3.77
GTI Faced North	0.90	16.86	8.15	-0.53
GTI Faced West	0.91	45.99	21.45	-4.86
GTI Faced East	0.91	47.97	21.28	-4.56
GTI Faced South	0.89	75.67	35.29	-3.16

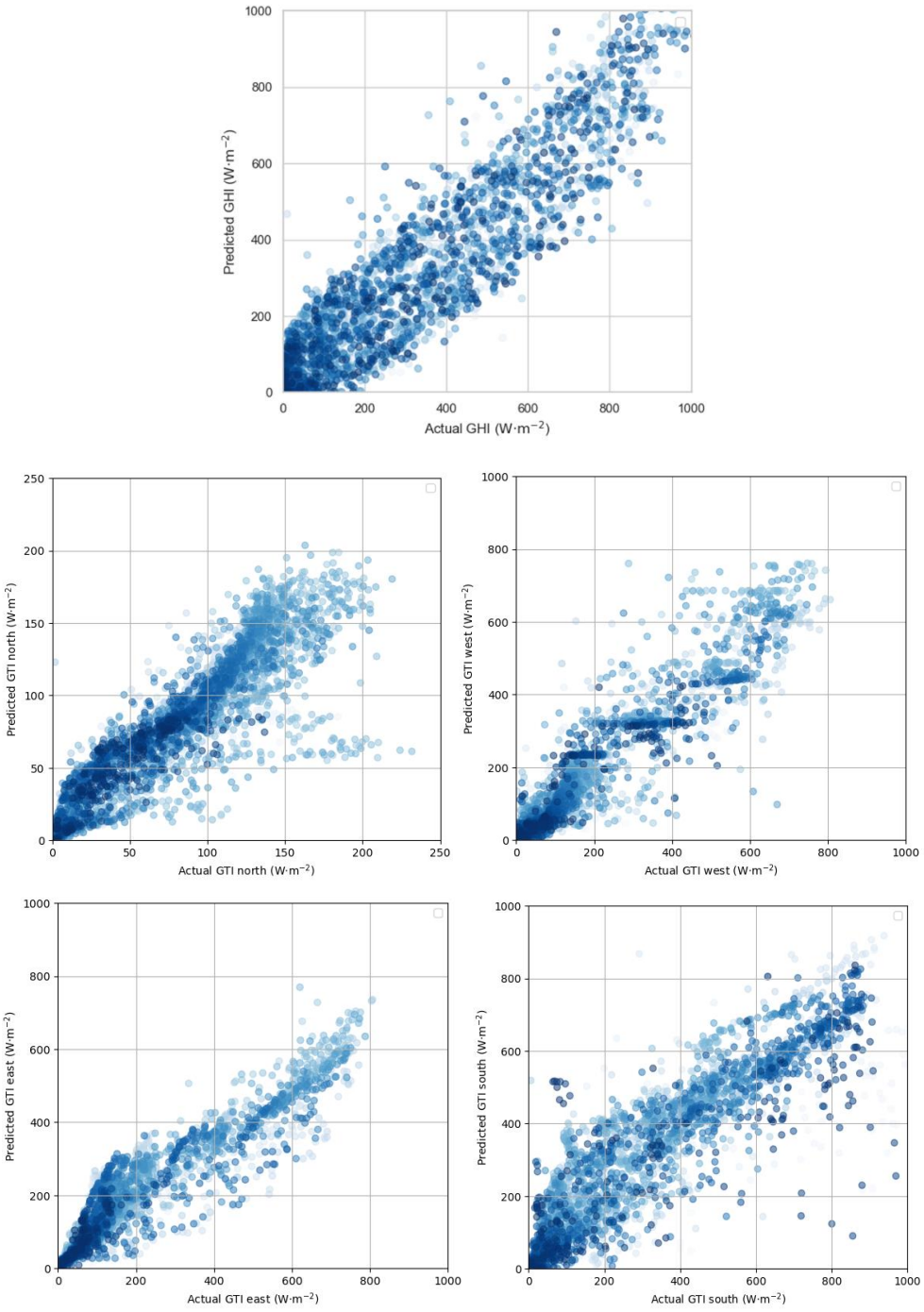

Figure 3: Comparative scatter plots of predicted vs. measured Global Horizontal (GHI) and Tilted (GTI) solar irradiance by cardinal orientations (testing dataset) at NOI Techpark weather station – assessing the novel RF-Symbolic Regressor model's accuracy

3.3 Resource utilization analysis for solar irradiance prediction

Modern approaches in symbolic regression aim to create models that efficiently generate mathematical expressions, employing optimization techniques, genetic algorithms, and other strategies to navigate the vast search space. These

efforts are crucial for making symbolic regression practical for problems where interpretability is vital despite computational challenges.

Table 2 analyzes resource utilization in our hybrid RF-Symbolic Regression model, assessing computational efficiency and practical viability. Peak memory usage varies across cardinal orientations, with *GTI* facing east requiring the most memory (1461.80 MB) and *GTI* facing north the least (340.35 MB), reflecting orientation-dependent complexity.

Execution times range from 35.41 s for *GTI* facing west to 172.89 s for *GTI* facing east, indicating varied processing complexity. Despite longer processing times for some orientations, all remain reasonable for real-time applications.

CPU utilization percentages range from 26.6% for *GTI* facing east to 37.6% for *GTI* facing west, indicating efficient resource harnessing. Overall, our model strikes a balance between precision and computational complexity, offering accurate predictions within reasonable resource limits for various applications.

Table 2: Resource utilization comparison for baseline and proposed approaches

Parameter	Maximum memory usage [MB]	Execution time [s]	CPU utilization [%]
GHI	419.30	111.49	32.1
GTI Faced North	340.35	103.72	29.0
GTI Faced West	344.24	35.41	37.6
GTI Faced East	1461.80	172.89	26.6
GTI Faced South	820.43	78.21	31.3

4. CONCLUSION

In conclusion, our study has introduced a novel Hybrid Random Forest-Symbolic Regression model to predict hourly solar irradiance, particularly in challenging environments such as mountainous and urban regions. The importance of our research holds relevance in the domains of renewable energy applications and environmental planning, providing valuable insights and practical implications. Our Hybrid model showcases a high accuracy, substantiated by high Coefficient of Determination (R^2) values ranging from 0.88 to 0.91 across various cardinal orientations. This robust performance underscores its proficiency in elucidating a significant portion of solar irradiance variability.

Moreover, our analysis of resource utilization during mathematical expression generation reaffirms the model's practicality. It consistently delivers precise solar irradiance predictions while remaining within reasonable resource consumption limits. Memory usage spans from 340.35 MB to 1461.80 MB, execution times range from 35.41 s to 172.89 s, and *CPU* utilization percentages vary between 26.6 % and 37.6 %.

Importantly, our model not only achieves accuracy but also generates interpretable mathematical expressions that capture intricate relationships between input variables and solar irradiance, even under challenging conditions involving shading and terrain effects. These expressions offer invaluable insights into the underlying physical processes, empowering better comprehension and decision-making.

Looking ahead, we aim to optimize our model's efficiency for real-time deployment and broaden its use across diverse geographical regions. Adding more meteorological data could enhance accuracy, while exploring hybrid models offers potential for precision. We're also interested in its application in urban planning for energy-efficient designs. In summary, our Hybrid Random Forest-Symbolic Regression model shows promise for accurate solar predictions, serving various fields such as renewable energy planning and environmental assessments. We're committed to advancing its capabilities to support the transition to sustainable energy sources.

REPRODUCIBILITY STATEMENT

The authors promote research reproducibility by providing their source codes upon request. This fosters transparency, allowing fellow researchers to verify their methods and algorithms in solar irradiance prediction. For access, please contact the corresponding author, encouraging scientific collaboration.

ACKNOWLEDGMENT

This research was funded by the internal project of the Free University of Bozen-Bolzano “SOMNE - Bolzano Solar Irradiance Monitoring Network” (CUP: I56C18000930005; CRC Call 2018).

REFERENCES

Abdellaoui, I. A., & Mehrkanoon, S. (2021). Symbolic regression for scientific discovery: an application to wind speed forecasting. *arXiv:2102.10570v2*. https://doi.org/10.48550/arXiv.2102.10570.

Banzhaf, W., Francone, F.D., Keller, R.E., & Nordin, P. (1998). *Genetic programming - An Introduction: On the Automatic Evolution of Computer Programs and Its Applications*. Morgan Kaufmann Publishers, Inc.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Solanki, S. (2022, August 5). *How to Use "memory_profiler" to Profile Memory Usage by Python Code?* CoderzColumn. https://coderzcolumn.com/tutorials/python/how-to-profile-memory-usage-in-python-using-memory-profiler

Eiben, A. E., & Smith, J. E. (2003). *Introduction to Evolutionary Computing*. Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-662-44874-8-

Python time module. (2023, November 21). GeeksforGeeks. Retrieved November 23, 2023, from https://www.geeksforgeeks.org/python-time-module/

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225-2236. https://doi.org/10.1016/j.patrec.2010.03.014.

Gonzalez Zelaya, C. V. (2019). Towards explaining the effects of data preprocessing on machine learning. *In 2019 IEEE 35th International Conference on Data Engineering (ICDE)* (pp. 2086-2090). IEEE. https://doi.org/10.1109/ICDE.2019.00245.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings, 3*(1), 91-99. https://doi.org/10.1016/j.gltp.2022.04.020.

Nugroho, H., Utama, N. P., & Surendro, K. (2021). Normalization and outlier removal in class center-based frefy algorithm for missing value imputation. *Journal of Big Data*, 8, 129. https://doi.org/10.1186/s40537-021-00518-7.

Patro, S. G. K., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. *arXiv:1503.06462v1*. https://doi.org/10.48550/arXiv.1503.06462.

Rodola, G. (2023, June 4). *Psutil documentation — psutil 5.9.6 documentation*. Psutil. https://psutil.readthedocs.io

Ruggiero, R. (2020, November 17). *Symbolic Regression: The Forgotten Machine Learning Method*. Towards Data Science. Retrieved November 23, 2023, from https://towardsdatascience.com/symbolic-regression-the-forgotten-machine-learning-method-ac50365a7d95

Schmidt, M., & Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324(5923), 81-85. https://doi.org/10.1126/science.1165893.

Solcast. (n.d.). *Solar API and Solar Weather Forecasting tool*. https://solcast.com/

Stephens, T. (2016). *Gplearn: Genetic Programming for Symbolic Regression*. GitHub. https://github.com/trevorstephens/gplearn

Bolzano - Bozen Topographic map, elevation, terrain. (n.d.). Topographic maps. https://en-gb.topographic-map.com/map-2513l/Bolzano-Bozen/

Udrescu, S.-M., & Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16). https://doi.org/10.1126/sciadv.aay2631.