Peer-Review

Kim, Hyeonwoo, and Hyeonjoon Lee. 2025. "Solar Energy Forecasting in Gwangju Using PM10 Data and Physics-Informed Neural Networks." *Journal of High School Science* 9 (2): 325–45.

The manuscript cannot be considered for publication in its current form. The following concerns need to be addressed to improve the quality and clarity of the work:

1. Comparison to Existing Models in Gwangju:

The city of Gwangju, South Korea, already employs modeling techniques similar to those described in the manuscript. The authors should clarify how their modeling approach compares to those used in Gwangju. What novel insights or improvements does their model offer?

2. Power Generation Context and Model Accuracy:

Gwangju's energy mix includes coal and LNG power plants, renewable sources (such as solar and wind), and biomass from municipal solid waste. Given this diversity, a rough modeling error margin of up to 20% might be considered acceptable. How does the proposed model's accuracy compare within this context?

3. Citations and References:

The manuscript lacks proper citations in several key areas. The authors must ensure that all relevant claims and data are appropriately referenced throughout the paper.

4. Data Source Transparency:

The authors should provide a clear and accessible link to the data sources used for model training and evaluation.

5. Baseline Comparison Using Simple Models:

Have the authors conducted any basic or back-of-the-envelope modeling using simple equations or statistical fits to compare with the AI-based results? If not, they should explain why this was not done. If such comparisons exist, the outcomes should be presented for context.

6. Expanded Introduction and Contextual Detail:

The introduction requires additional context. For instance, what are the typical outcomes when no modeling is used, or when existing methods are applied? How might this affect energy planning or operations in the region? Quantitative details such as the proportion of solar power in the total energy mix of Gwangju would help frame the study's relevance (see also point #2).

7. Clarification of PM Impact:

The statement that "our model can learn that hazy days with high PM lead to lower PV output" oversimplifies a known phenomenon. The authors should clarify what specific role PM10 plays in this relationship beyond what is already expected due to reduced sunlight during hazy conditions.

8. Model Usability and Real-World Application:

More information is needed about how the model is intended to be used operationally. For example:

◦ Does an operator input the eight parameters manually?
◦ What happens if a parameter is missing or the forecast is inaccurate?
◦ How sensitive is the model to errors in input data?
◦ Is the model intended for real-time use?

Addressing these questions will help readers assess the model's practical utility.

9. Predictive Use for Missing Parameters:

If the model performs well, can it be used to infer or predict unknown parameters (e.g., PM10) when those are not directly available? Is PM10 a sensitive enough feature to support this?

10. Low Correlation Justification:

The reported correlation between PM10 and solar output ($r \approx -0.2$) is relatively weak. The authors should justify the inclusion of PM10 as a feature and discuss why other potentially relevant parameters (e.g., PM2.5, wind speed, wind direction) were not considered. Please refer to literature such as "Understanding Solar Irradiance Predictions Through Explainable AI: Perspectives from Gwangju" by

Yang et al., which explores these aspects.
11. Issues with Figure Presentation (Model Predictions vs. Actual):
a) The Y-axis lacks units.
b) The maximum value on the Y-axis (~40 kW) appears inconsistent with the stated expected output (~866 kW).
c) The clarity of the figure is poor; it is difficult to distinguish between the different models and the actual data. Improving visualization would enhance interpretability.
12. Insignificant Differences in Model Performance:
The differences in MSE between the models are quite small. When considering the root mean square error (RMSE), the variations are negligible. The authors should justify their claims of model accuracy improvements. Additionally, what level of prediction error is acceptable for operational use by utilities? How does the model's error (e.g., ~5 kW) compare to what would be useful in practice? (See also points #1 and #2.)
13. Interpretation of PM10 Impact on Model Performance:
The statement that including PM10 improved model performance by reducing MSE by ~10% should be carefully evaluated. A 10% change in MSE may not be practically significant in this context. The authors should provide a more nuanced interpretation of these results (see also points #10 and #12).
Overall Recommendation:
The topic is of interest, and the approach has potential. However, the manuscript needs significant revision to improve clarity, completeness, and justification of results. I encourage the authors to address the above points comprehensively to enhance the quality and impact of their work.

Dear Editor,
We are submitting the revised version of our submission for the paper. First of all, we would like to thank you and the reviewer for your pertinent and helpful comments. We addressed all the suggestions to clarify our contribution and all the methodological and experimental aspects that were a matter of question by the reviewer. Also, we checked the grammatical errors and revised the manuscript in a more articulated manner. In this new version of the manuscript, you will find in yellow highlighted changes that have been made to address the reviewer's observations. The responses to the report and the reviewer's suggestions are detailed hereafter.

Responses to Reviewer

Comment 1: **Include a comparison with existing models used in Gwangju, clearly explaining the novelty or improvements of the proposed model.**
Reply:
We appreciate this suggestion. In response, we have added a discussion of existing solar forecasting models in Gwangju/Korea and highlighted how our approach differs. For example, we now mention prior studies using XGBoost and LightGBM for irradiance prediction in Gwangju and recent deep learning models across Korean regions. We explain that unlike these purely data-driven models, our work introduces a physics-informed neural network that integrates domain knowledge (solar geometry and atmospheric attenuation) and includes PM10 data. This clearly positions the novelty of our model: it bridges the gap between physics-based and machine learning approaches and incorporates air quality effects that others have not.

Comment 2: **Contextualize model accuracy within Gwangju's diverse energy mix and justify the model's error margin.**
Reply:

We have expanded the discussion to put our forecasting accuracy in context. Specifically, we note that South Korea's energy mix is dominated by nuclear and fossil fuels with only ~6–9% from solar/wind. Therefore, a 10% error in solar forecasts currently has a minor impact on total grid supply (approximately 0.6% of total generation). We justify that our model's error (RMSE ~10–12% of peak output) is within acceptable operational margins given this mix. We also add that as Gwangju and Korea aim for higher renewable shares (e.g., 20% by 2030), maintaining or improving forecast accuracy will be critical. In essence, we assure that the current error margin is justified and manageable, and we emphasize the model's value as solar penetration grows.

Comment 3: Add missing citations and references where needed.
Reply:
We have carefully reviewed the manuscript and inserted appropriate citations for all key statements.

Comment 4: Provide a clear link and explanation of data source transparency for model training and evaluation.
Reply:
We have added a dedicated subsection ("Data Sources and Transparency") under Methodology detailing all data used. We explicitly state the sources.

Comment 5: Compare results with simple models (e.g., statistical fits or basic equations) and explain their role.
Reply:
We appreicate the reviewer's comment. Nonetheless, we have not found a comparative model in this context of tasks, thus we defined the simple MLP ourselves and made a comparison.

Comment 6: Expand the introduction with more regional context: e.g., Gwangju's solar power proportion, outcomes without modeling, and how this aids planning.
Reply:
We thanks the reviewer's point. Along with the Comment 1, we have significantly expanded the Introduction to include regional context accordingly.

Comment 7: Clarify and expand on the impact of PM10, avoiding oversimplification.
Reply:
We have expanded our discussion on how PM10 affects solar radiation and thus forecasting. We explicitly state that the effect is complex: while day-to-day linear correlation is low, PM10 can still cause significant attenuation under high pollution conditions. We included literature evidence that even a small increase in PM10 leads to measurable drops in PV output. We describe how our model captures these effects (improving forecasts on hazy days) and avoid any suggestion that PM10 is the dominant factor at all times. By providing a nuanced explanation – mentioning direct attenuation, soiling, and conditional importance of PM10 – we ensure the impact is not oversimplified. We also clarified that on clear days PM10 has minimal effect, whereas on polluted days it's important, aligning with physical reality.

Comment 8: Detail the model's operational usage: parameter input methods, handling missing/inaccurate inputs, sensitivity, and real-time applicability.
Reply:

We highly appreciated the reviewer's comment. Nonetheless, we believe the real-time applicability including the deployment scenario of the model is not the main focus of the paper. We will attach the appendix part regarding your comment after the acceptance.

Comment 9: Discuss the possibility and limits of inferring unknown parameters (e.g., PM10) using the model.
Reply:
We addressed this in the Discussion. We acknowledge that one might try to invert the model to estimate PM10 from known outputs, but we clarify the limitations. We explain that while a large unexplained drop in output could hint at high PM10, multiple factors can cause similar drops, so the inference is not unique.

Comment 10: Justify including PM10 despite low correlation, and discuss other factors like PM2.5, wind, with literature support.
Reply:
We have strengthened the justification for including PM10. We explain that low overall correlation does not mean PM10 is useless – citing evidence that aerosols significantly reduce solar output in certain scenarios.

Comment 11: Improve figure presentation: add units to axes, correct inconsistencies in value ranges, and enhance clarity.
Reply:
For better clarity, we added the units in the text and explained the figure in more detailed tone.

Comment 12: Evaluate the significance of performance differences (e.g., MSE vs. RMSE), and clarify operational relevance of errors
Reply:
We have expanded our discuussion to address this. We added sentences in the Discussion ("Contextualizing Accuracy") about what an error means for grid operations.

Comment 13: Offer a nuanced explanation of the ~10% MSE improvement due to PM10 inclusion.
Reply:
We have provided a more detailed explanation for the ~10% MSE improvement observed. We explicitly describe how the improvement manifests (mainly during high PM periods) and what it implies.

The reviewer appreciates the authors' efforts in addressing the previous comments. However, two points remain partially unresolved:
Regarding point #8: The authors are kindly reminded that, if the manuscript is accepted for publication, it is expected that all relevant code and data files be made publicly available (e.g., via GitHub or a similar platform), along with clear instructions for their use.
Regarding comment #11 on Figure 1: The Y-axis still lacks units, and the discrepancy between the maximum Y-axis value (~40 kW) and the stated expected output (~866 kW) has not yet been adequately explained.

Dear Mr Apte
First of all, thank you for thoughtfully looking through our paper and providing us with feedback. We believe that this iteration of submission -> feedback is absolutely necessary for ensuring that the paper is up to standards for publication, hence, we would like to express a deep gratitude for your effort and for giving us a chance to improve our manuscript.
This is a comment response letter that'll explain detailedly how we have responded to the two (point 8 and point 11) points that Mr Apte has pointed out and what changes we have made to the manuscript in response.

Comment 8 (The authors are kindly reminded that, if the manuscript is accepted for publication, it is expected that all relevant code and data files be made publicly available (e.g., via GitHub or a similar platform), along with clear instructions for their use.)
Response for comment 8: In response to the comment, we've removed the "Code Release" section in the manuscript and added an URL of the github repository of our work in the abstract. We also removed the following from the abstract: " The full dataset and implementation will be released publicly upon publication to encourage further research in physics-informed solar modeling." The github repository will also be available on this comment response letter and, it contains the actual code that generated the model and trained it, the actual dataset that we've used, and the code that we've used to double check the maximum value of our training and test dataset <- this was done for comment 11.

Comment 11 (on Figure 1: The Y-axis still lacks units, and the discrepancy between the maximum Y-axis value (~40 kW) and the stated expected output (~866 kW) has not yet been adequately explained.)
Response for comment 11: In order to address the lack of unit to the Y axis of Figure 1, we have added a text at the bottom of figure 1 that says "*Y axis unit: kW (Kilowatt)"* in order to indicate the unit of the Y axis. This change has been highlighted in blue in the document of the manuscript with highlights on changes.
Additionally, Mr. Apte has observed a discrepancy between the stated expected output (~866 kW) and the maximum Y-axis value (~40 kW). We would like to clarify that the ~866 kW figure refers to the maximum value of the target variable in the training dataset used for model development, not the expected output in the final plot. This confusion likely arose due to our earlier labeling, which lacked sufficient clarity. To address this, we have updated the label from *Range (Train)* to *Range (Training Dataset)* to provide greater precision. This revision has been highlighted in blue in the manuscript document with highlights on changes.
Moreover, we have used a max() function on jupyter notebook to double check the range of the target variable of our training dataset and test dataset. And we've found that the actual maximumvalue for the training dataset was exactly 49.526795. Therefore, we've also made an update to the range of our target variable in the training dataset in the manuscript from (*0kW to ~866kW)* to *(0kW to 49.526795kW).* This update also has been highlighted in blue in the manuscript document with highlights on changes.

Additional Notes regarding comment 11:
When we ran a code on a jupyter notebook to check the maximum value of the target variable for each datasets, the maximum value of the target variable that we got for the **Test Dataset** was exactly 49.867033. And if you anticipate on figure 1, you'll see that the peak isn't around 50 and is rather much closer to 40 (~40kW) as you've mentioned. This is because the peak occurs at sample number 5733, but our plot only shows the first 100 samples. This has been done in this way because, if we tried to draw all test points at once, the lines would overlap so densely that

the chart would look like a solid block instead of a readable graph. The purpose of the graph for this paper was to provide a focused view of model versus actual values on a representative subset of the test data, highlighting key trends without clutter, while overall performance is highlighted by the MSEs that were calculated using the entire test dataset.

Since we believe that it's our responsibility to make the paper as understandable as possible, We identified the exact sample index at which the peak occurs and determined the corresponding peak value via jupyter notebook, and we've found that the peak value occurs at sampe number 24 with a corresponding value of exactly 42.8792 (42.8792 kW). This information has been added into the revised manuscript, highlighted in blue, and the code that we've used to identify this is also available in our Github repository.

Final words from the author:
Again, thank you for reviewing our work, and have a nice day!

Link to the Github Repository:

https://github.com/TheGitpilot/Solar-Energy-Forecasting-in-Gwangju-Using-PM10-Data-a nd-Physics-Informed-Neural-Networks/tree/main

---

Please have your manuscript confirm to the Journal's formating requirements. In particular: Please provide a word document of your manuscript. 12 Times Roman font, single column. Make sure referenes follow a consistent format. If there are more than 6 authors per reference, all the 6 need to be listed followed by an et al., if there are < 6 authors, all 6 need to be listed. The Journal name, volume, issue, pages, year need to be listed followed by a live link (DOI). The same format needs to be followed for each reference. Please renumber the references manually (do not use the word software for numbering). All tables need to be inserted as Tables (not separate lines of text separated by spaces). Please make sure figures are of adequate resolution and delete Figure titles and legends from figures (submit separately or as part of the word text). The X and Y titles need to incorporate units as part of the figure.

Dear Reviewer,

First of all, we would like to express a deep gratitude for your comments. I'm happy that you were satisfied with our responses. Again, Thank you.

For an additional requirement, the reviewer has asked us to do the following:
- Provide a word document
- Remove the legend and title from the graph, and add units
- Use Tables
- Manually number the references, keep consistent style, and add a DOI link

In response, we've provided a word document, with editing on graphs, insertion of tables, and alternation of references in harvard style followed by a DOI link.

Thank you for reviewing our work, and have a nice day! From the author.

Accepted.