I did not see how the EPI was actually calculated. You have 3 levels and 10 factors, which means that you need to perform 3^10 sets of experiments (or need 3^10 sets of data, each with its own calculation using multivariate analysis with 10 variables (assuming no collinearity, heteroskedasticity and equivalence of variances for each factor). I don't see how this is remotely possible.
The only reason I give this manuscript a revise and resubmit instead of a reject is that I may have misunderstood the premise of the article.
Please present how the actual 3^10 sets of data were obtained and how possibly could each data set have been calculated. Present regression equations and EPI calculations in depth.

To,
The Editor,
Journal of High School Science

03rd Sept 2024

Subject : Revised version of our manuscript titled "Leveraging Surrogate Modeling using synthetic data for Energy Efficient Decision-Making in Residential buildings"

Respected Sir,
We sincerely appreciate your feedback on our paper titled *'Leveraging Surrogate Modeling Using Synthetic Data for Energy-Efficient Decision-Making in Residential Buildings.'* In response to the reviewer's comments, we have revised the paper, adding explanations and additional information to address the concerns raised. The revision in sections 2 and 2.1 are highlighted in red text on pages 4, 6, and 7 of the revised manuscript. Below, we provide our detailed response to the clarifications requested by the reviewer.

Reviewer Comments

I did not see how the EPI was actually calculated. You have 3 levels and 10 factors, which means that you need to perform 3^10 sets of experiments (or need 3^10 sets of data, each with its own calculation using multivariate analysis with 10 variables (assuming no collinearity, heteroskedasticity and equivalence of variances for each factor). I don't see how this is remotely possible.
The only reason I give this manuscript a revise and resubmit instead of a reject is that I may have misunderstood the premise of the article.
Please present how the actual 3^10 sets of data were obtained and how possibly could each data set have been calculated. Present regression equations and EPI calculations in depth.

Response

A 3D building thermal modeling approach was adopted to generate a synthetic dataset, which was then used to train and develop an ANN-based surrogate model. We then employed classical algorithms and Artificial Neural Networks (ANNs), rather than exhaustive combinatorial methods, to train models on this synthetic dataset. This allowed the models to learn the complex, non-linear relationships between the 10 input factors and the EPI. The models generalize well to new, unseen data, making the process both efficient and scalable. Optimization was performed using gradient descent to efficiently identify the best model parameters without manually computing every possible scenario.

Building energy simulation models are sophisticated tools (eQuest, Energyplus, DesignBuilder, Openstudio etc) that simulate buildings and estimate their EPI. The process begins with the creation of a 3D model of the building, incorporating various inputs that influence energy consumption, such as heating, cooling, lighting, ventilation, and other operational needs. Key inputs include building geometry, orientation, construction materials, envelope characteristics, internal loads, occupancy patterns, lighting and equipment loads, HVAC system types, and standard weather data.

The simulation model calculates thermal loads, and the operation of air conditioning, lighting, and equipment based on usage patterns, hourly weather data, and control mechanisms. It also considers interactions between different systems; for example, heat generated by lighting and equipment may reduce heating loads in winter but increase cooling demands in summer. These interactions are crucial for accurate energy consumption estimates.

The simulation is conducted on an hourly basis over an entire year, calculating energy consumption for each hour while accounting for dynamic interactions between building systems and the external environment. A thermal simulation model in EnergyPlus was developed, and scripting was used to modify one parameter before executing each simulation. The resulting EPI values were extracted into a CSV file for analysis. Using this scripting approach, we generated and simulated 116,640 parametric runs. An extract from the dataset generated through parametric building thermal simulation modelling, illustrating key input variables and corresponding Energy Performance Index (EPI) outputs and the approach has been added to the research paper for better understanding.

I hope that you will find our revised manuscript in order. The revised version is submitted for publication in your esteemed journal.

Thanking you,

Yours Sincerely
Mannat Kaur

Thank you for addressing my comments. See additional concerns based on your comments below. Please address these concerns in the manuscript.
1.What you have done is hence to enter data into a pre-constructed simulation model available in the public domain; then enter the results from that simulation into various other ML algorithms and models, again available in the public domain; and finally, reported the results from those simulations and models. I submit that these are data-entry operations and do not require the use of scientific acumen. Therefore, it is questionable whether this manuscript may be construed as a scientific body of work.

2.Your model is rigid and can only work with your parameters. For example, if I was to build a house facing north-west, with a window to wall ratio of 55%, 30% shaded, an RCC slab roof with 75 mm insulation…….. I would need to data enter an entirely new set of (synthetic) output EPI values (generated by the algorithm) into the ML to find the optimum model again. With this different data set, a different ML model may perform better. Hence, your results are limited and incapable of generalization, partly because it uses ordinal/categorical (discrete) values for many parameters.

3.Even if you did enter parameters with values from point 2 as test data into your model, there would be no way to know if the EPI calculated was actually correct for that configuration, unless you input those values into the simulation (which defeats the purpose of the ML model). I would like to see data (at least 10% of training) that has different values than any of your categorical parameters (see point2), entered into your best trained model and then see how it performs. You will need to enter the data in the simulation as well to obtain exact EPI values to compare against. Please present in the manuscript.

4.Going back to point 3, is there any use in inputing this data into a ML model at all ? If all you needed was an EPI, you could enter whatever values you wanted into the simulation and get yourself an EPI number.

5.You state that "…This approach provides insights into which specific building features most significantly impact the building's Energy Performance Index….." How and what exactly are these insights? How are they interpreted from the model's output data? And will these insights not be dis-similar depending upon input data (see point 2). For example, for data with different input ordinal/categorical or discrete parameters, a different model may perform better, in which case; the insights obtained may be different. (for example, see feature importance rankings with different models)

To,
The Editor,
Journal of High School Science

18th Sept 2024

Subject: Revised version of our manuscript titled "Leveraging Surrogate Modeling using synthetic data for Energy Efficient Decision-Making in Residential buildings"

Respected Sir,

We sincerely appreciate your critical and highly constructive feedback on our paper titled *'Leveraging Surrogate Modeling Using Synthetic Data for Energy-Efficient Decision-Making in Residential Buildings.'* In response to the reviewer's comments, we have revised the paper, adding explanations and additional information to address the concerns raised. The revision in sections 4, section 4.3 are highlighted in red text on pages 23 and 24 of the revised manuscript. Below, we provide our detailed response to the clarifications requested by the reviewer.

1)What you have done is hence to enter data into a pre-constructed simulation model available in the public domain; then enter the results from that simulation into various other ML algorithms and models, again available in the public domain; and finally, reported the results from those simulations and models. I submit that these are data-entry operations and do not require the use of scientific acumen. Therefore, it is questionable whether this manuscript may be construed as a scientific body of work.

Response:

The body of work presented in this paper involves several interconnected tasks, with synthetic dataset generation and input being just one part of the process. The research primarily focuses on evaluating the performance of various ML models to determine which is most suitable for building industry applications. As different machine learning models can vary significantly in accuracy and performance across application domains, this evaluation is crucial. These variations depend heavily on data characteristics such as structure, distribution, and the relationships between features. For instance, a model that performs well on data with more linear relationships might struggle with data involving non-linear or complex interactions. Additionally, the development and training of the Artificial Neural Network (ANN) model, in particular, required significant technical input such as hyperparameter tuning, optimizing network architecture, adjusting learning rates and regularization techniques.

To summarize, the paper focuses on three primary contributions:

1. Model Suitability and Performance:

Assess and evaluate the suitability, performance, and precision of ML models (both classical and neural networks) trained on data generated through building thermal simulations. Our research demonstrates the effectiveness of these models in predicting building Energy Performance Index (EPI) using generated simulation data.

2. Reduced Computational Load:
A key contribution is the significant reduction in computation time required for energy simulations. On average, preparing a simulation takes around 4-5 days and a single simulation run takes about 2-5 minutes. During the early design stages, design teams must simulate energy consumption across numerous combinations of input parameters to arrive at the most optimized and sustainable design. Running simulations for all these combinations requires immense computational resources, which can be mitigated by using pre-trained machine learning models. For example, simulating energy consumption for five parameters with ten variations each results in 100,000 combinations. This also means that the simulations can now be run on thin-clients, such as cellphones without having to deal with heavy infrastructure.

3. Overcoming Early-Stage Design Challenges:
Small residential projects often lack access to experienced energy modelers during the early design stages, when the design has the greatest flexibility to incorporate energy-saving measures. A pre-trained, quick ML model can drive sustainable, low-carbon design decisions early in the process, offering maximum benefits.

2)Your model is rigid and can only work with your parameters. For example, if I was to build a house facing north-west, with a window to wall ratio of 55%, 30% shaded, an RCC slab roof with 75 mm insulation…….. I would need to data enter an entirely new set of (synthetic) output EPI values (generated by the algorithm) into the ML to find the optimum model again. With this different data set, a different ML model may perform better. Hence, your results are limited and incapable of generalization, partly because it uses ordinal/categorical (discrete) values for many parameters.

Response:
As outlined in the abstract and paper, the primary focus was to develop a tool that supports better decision-making during the early design stages of residential projects. In the early stages of building design, the available information is often broad and conceptual rather than highly detailed. At this point, architects and design teams typically work with generalized parameters such as approximate window-to-wall ratios, rough building massing, and climate zone data. Specific details, such as exact window placements, shading dimensions, or precise material choices, are usually unavailable. This stage focuses on exploring design concepts and understanding how major factors like window sizes, wall materials, and orientation affect overall performance. It is only during the detailed design phase that more granular information—such as exact WWR percentages, shading profiles, and refined material specifications—becomes available for precise modeling.
So, the main utility of this approach lies in the early design stage, where it's often challenging to get an energy modeler on board. At this phase, quick decision-making tools are crucial to guide sustainable design choices. However, as suggested in pt3, we have tested the performance of ANN model on interpolated values to check the generalization and it showed positive results. We have added the same in Section 4.3 of the paper.

3)Even if you did enter parameters with values from point 2 as test data into your model, there would be no way to know if the EPI calculated was actually correct for that configuration, unless

you input those values into the simulation (which defeats the purpose of the ML model). I would like to see data (at least 10% of training) that has different values than any of your categorical parameters (see point2), entered into your best trained model and then see how it performs. You will need to enter the data in the simulation as well to obtain exact EPI values to compare against. Please present in the manuscript.

Response:
We did perform a train-test split during our machine learning process - here, we set aside a random data subset and trained on the rest. Then, we tested it on the unseen random sample; many of these would indeed have totally unseen parameters. There is definitely some danger in terms of overfitting to previously seen data and losing the ability to generalize, but we have used standard processes for hyperparameter tuning to make sure that this does not happen.
Also, as suggested we did additional simulation models (~ 10% of original dataset) to check how the model performs on a completely unseen dataset. The results are positive and the information for the same has been added in section 4.3 of the paper.

4)Going back to point 3, is there any use in inputting this data into a ML model at all? If all you needed was an EPI, you could enter whatever values you wanted into the simulation and get yourself an EPI number.

Response:
In addition to the points mentioned in response to Pt 1, following are few reasons to use ML in these scenarios:
Where running the simulator is too expensive (computationally and time-wise) - e.g., on a thin-client, such as a cellphone. Most modern cell phones have excellent ML capabilities and will allow for an on-the-spot and immediate estimate.
Longer term insights in terms of feature importance (we mention this, but do not go into much detail on the topic in this paper).
The ML model (especially the ANN) provides a reasonable accuracy alternative for people who cannot purchase expensive simulation software. Further, extended use of the software in a particular area (i.e., the data produced in-house by its usage within a company) can be used to fine-tune the system over time.

5)You state that "…This approach provides insights into which specific building features most significantly impact the building's Energy Performance Index….." How and what exactly are these insights? How are they interpreted from the model's output data? And will these insights not be dis-similar depending upon input data (see point 2). For example, for data with different input ordinal/categorical or discrete parameters, a different model may perform better, in which case; the insights obtained may be different. (for example, see feature importance rankings with different models)

Response:
Our exploration of different models reveals which features are likely to be most important when computing the EPI (which may not otherwise be obvious, given the large number of variables and complex interactions involved).
Feature significance may indeed change given different data parameters and model choices, and this is useful: for example, XGBoost is known to ignore highly correlated features (So if a, b, c, d are are your features and a,b are highly correlated, then Xgboost model may randomly choose one - say a - at this

point it realizes that choosing b as an important feature also is useless! This is because it already has a, so adding b will not give the model more power.)
Background: Random forest uses bagging ensemble model while XGBoost uses boosting ensemble model. XGBoost calculates feature importance based on metrics like Gain (improvement in predictive power), Cover, and Frequency (how often a feature is used). Random Forest, on the other hand, measures feature importance based on the decrease in node impurity (Gini or entropy) across all trees where a feature is used for splitting.

I hope that you will find our revised manuscript in order. The revised version is submitted for publication in your esteemed journal.

Thanking you,

Yours Sincerely
Mannat Kaur

Thank you for addressing my comments. However,

1.Did you test with 1164 runs? That will be 1% of your data set (even though I mentioned 10% of training in my previous review, I think that may be too many runs, I will settle for 1%, which is 1164 rows). You do not mention how many runs in section 4.3. Also, you do not present a feature importance historgram for ANN (your chosen method). Please present this in the manuscript. There is not much point to changing COP (which is the most important feature in XGboost) if it is also not the most important feature in ANN. Hence, I want you to change the three most important features of ANN to values that interpolate between 33% and 66% of their fed values, one at a time, for different fed values of the other parameters and present the calculated EPI versus the actual EPI for those 1164 runs. You mention that changing the COP values did not change the MAPE by a significant amount; on the contrary, now your model only explains 80% of the variance and your MAPE is two-orders-of magnitude greater.

2.When you present this data and analysis, please present average difference between EPI for the calculated and predicted values as well as the standard deviation (goes back to my point 3 from my earlier review).

3.Regarding your response to point 1, the simulation is free of cost. Building a house is not that urgent a task that needs to be performed in the absence of resource competent hardware or software (such as a phone). Therefore, I do not buy into your 1 and 2 responses. However, you can put down response 3 in the manuscript.

4.In response to my earlier comment you state that "…..Our exploration of different models reveals which features are likely to be most important when computing the EPI (which may not otherwise be obvious, given the large number of variables and complex interactions involved)….." However, I find that your features vary significantly depending on which model is chosen. For example, climate, BC and glass are the most important in the RF model whereas COP, shade and WWR are most important in XGboost. You do not have a histogram for ANN (see point 1). Hence, notwithstanding the reasons why there are drastically different feature importances for different models (you have not performed a recursive feature elimination, nor have you performed a permutation feature imporatance or a correlogram or a heat map or tested for multicollinearity……) , this goes back to my earlier comment # 5, where these "insights" seem dis-similar and depend on input data and the type of model you choose.

Please provide significantly more depth to this manuscript in terms of data manipulation, methods, analysis, content and erudition; rather than responding only piecemeal to comments.

To,
The Editor,
Journal of High School Science

29th Sept 2024

Subject : Revised version of our manuscript titled "Leveraging Surrogate Modeling using synthetic data for Energy Efficient Decision-Making in Residential buildings"

Respected Sir,

We sincerely appreciate your critical and highly constructive feedback on our paper titled *'Leveraging Surrogate Modeling Using Synthetic Data for Energy-Efficient Decision-Making in Residential Buildings.'* We have carefully addressed the comments and incorporated the suggestions provided by the reviewer. Specifically, we have converted the categorical values in the original dataset into numeric values, which allows for better assessment of the model's generalization capabilities. As mentioned in our earlier response, the initial scope is to use this model in the early design phase, focusing on broad design options. However, we appreciate the importance of evaluating the model's performance on more granular, intermediate values, and this adjustment should enhance the model's applicability in that regard.

.

1) Did you test with 1164 runs? That will be 1% of your data set (even though I mentioned 10% of training in my previous review, I think that may be too many runs, I will settle for 1%, which is 1164 rows). You do not mention how many runs in section 4.3. Also, you do not present a feature importance histogram for ANN (your chosen method). Please present this in the manuscript. There is not much point to changing COP (which is the most important feature in XGboost) if it is also not the most important feature in ANN. Hence, I want you to change the three most important features of ANN to values that interpolate between 33% and 66% of their fed values, one at a time, for different fed values of the other parameters and present the calculated EPI versus the actual EPI for those 1164 runs. You mention that changing the COP values did not change the MAPE by a significant amount; on the contrary, now your model only explains 80% of the variance and your MAPE is two-orders-of magnitude greater.

Response:
In the previous iteration, the 5832 test runs were conducted for a single climate zone. Based on the valuable suggestion of the reviewer, we have now revised our approach and incorporated more detailed information, including the number of runs and the most important features, which is presented in Section 4.3.

2) When you present this data and analysis, please present average difference between EPI for the calculated and predicted values as well as the standard deviation (goes back to my point 3 from my earlier review).

Response: The manuscript has been updated to include the average percentage difference and standard deviation between the simulated EPI and the predicted values.

3) Regarding your response to point 1, the simulation is free of cost. Building a house is not that urgent a task that needs to be performed in the absence of resource competent hardware or

software (such as a phone). Therefore, I do not buy into your 1 and 2 responses. However, you can put down response 3 in the manuscript.

Response:
The suggestion is well taken, and we have included this aspect in the manuscript

4) In response to my earlier comment you state that "…..Our exploration of different models reveals which features are likely to be most important when computing the EPI (which may not otherwise be obvious, given the large number of variables and complex interactions involved) ….." However, I find that your features vary significantly depending on which model is chosen. For example, climate, BC and glass are the most important in the RF model whereas COP, shade and WWR are most important in XGboost. You do not have a histogram for ANN (see point 1). Hence, notwithstanding the reasons why there are drastically different feature importances for different models (you have not performed a recursive feature elimination, nor have you performed a permutation feature importance or a correlogram or a heat map or tested for multicollinearity……) , this goes back to my earlier comment # 5, where these "insights" seem dis-similar and depend on input data and the type of model you choose

Response:
Thank you for the insightful comment. We agree that the variation in feature importance across different models could create confusion for readers, as the results highlight different features depending on the model used. To address this and enhance clarity, we have included a graph for the best-performing model based on permutation feature importance in this version of the manuscript, as you suggested. This addition will provide readers with more consistent insights into which features are most influential based on the most reliable model. We acknowledge that different models yield varying feature importance rankings due to the inherent differences in their algorithms, and this graph should help readers gain a clearer understanding.

The comments provided by the reviewer have a valuable viewpoint and have improved our manuscript considerably. We hope you find the revised manuscript in order submitted for publication in your esteemed journal

Thanking you,

Yours Sincerely
Mannat Kaur

Thank you for addressing my comments. I have some more comments related to Figure 8 and the introduction of a table below figure 8 in the manuscript.

1.Check figure 8. You have 10 features but you report only 8 in the figure. You are missing building configuration and natural ventilation. Also, I do not understand how you only have 2 blue circles and one red circle (total 3) for wall type, when wall type has 4 levels in your table. similarly for roof type (that has 3 levels), but you only have 2 blue circles. Should the red circle not be in addition to the # of levels already in the matrix. I am also not able to understand the matrix of numbers above the red circles. Please explain this combination in the manuscript.

2.Should you not have a total of 810 runs? If wall type, roof type, orientation and AC are constants (intermediate values), you are changing 6 factors (glass, 3 levels), (w-w-ratio, 3 levels), (shade, 3 levels), (BC, 3 levels), (natural ventilation, 2 levels) and (climate 5 levels). = 810 runs. Please explain in detail in the manuscript.

3.Lastly, I would like to see a table similar to the "extract of the generated dataset" but with 2 more columns. Those columns should show the exact calculated value from the EnergyPlus simulation and the calculated value from the ANN. For rows 0, 1, 2, change Wall U value from 2.256 to 1.112 (everything else same), for rows 3, 4, 5, change Roof U value from 2.865 to 0.869 (everything else same), for rows 6, 7,8, change Orientation from 180 to 45 (everything else same) and from rows 9 and 10, change COP from 2.4 and 3 to 2.7 (everything else the same). As mentioned before, add two more columins and present the actual calculated value ffrom theEnergyplus in one column and the calculated ANN value in the other column. The objective is for the reader to double check their calculations against your so that your work can be replicated easily. This table can be presented after Figure 8. in the manuscript.

To,
The Editor,
Journal of High School Science

5th October 2024

Subject : Revised version of our manuscript titled "Leveraging Surrogate Modeling using synthetic data for Energy Efficient Decision-Making in Residential buildings"

Respected Sir,

We your feedback on our paper titled 'Leveraging Surrogate Modeling Using Synthetic Data for Energy-Efficient Decision-Making in Residential Buildings.' We have carefully addressed the comments and incorporated the suggestions provided by the reviewer. As suggested we have added more information in Section 4.3..

1) Check figure 8. You have 10 features but you report only 8 in the figure. You are missing building configuration and natural ventilation. Also, I do not understand how you only have 2 blue circles and one red circle (total 3) for wall type, when wall type has 4 levels in your table. similarly for roof type (that has 3 levels), but you only have 2 blue circles. Should the red circle not be in addition to the # of levels already in the matrix. I am also not able to understand the matrix of numbers above the red circles. Please explain this combination in the manuscript.

Response: As suggested earlier, we took steps to reduce the computational and simulation burden while still capturing the variation across key parameters. To achieve this, we limited categorical parameters like ventilation strategies and building configurations to just one type each. Figure 8 has been updated with ventilation strategies and building configuration for greater clarity.

Also, we took 2 values from the original dataset to ensure that the simulation and results were manageable and still representative, allowing us to effectively validate the ML model. So, we first modeled 1 unseen wall type along with 2 roof types, 2 orientations, and 2 AC performance levels from the original dataset. Other variables such as glazing types, Window-to-Wall Ratio (WWR), shading elements, and climate zones remained consistent with the original setup. This resulted in a total of 1,080 parametric runs.

Next, we adjusted the simulation to include 1 unseen roof type, while keeping the same known (2 values) wall, orientations and AC performance levels, which also produced 1,080 runs. We repeated this approach for the orientation and AC performance cases, which brought the total number of parametric runs to 4,320 (1,080 runs per case, multiplied by 4 variations).

The detailed explanation is also added in the section 4.3, please review the same.

2) Should you not have a total of 810 runs? If wall type, roof type, orientation and AC are constants (intermediate values), you are changing 6 factors (glass, 3 levels), (w-w-ratio, 3 levels), (shade, 3 levels), (BC, 3 levels), (natural ventilation, 2 levels) and (climate 5 levels). = 810 runs. Please explain in detail in the manuscript.

Response: A detailed explanation has been added on how we have arrived at 1,080 cases for each of the four-parameters resulting in total of 4,320 cases in Section 4.3. Please review the same.

3) Lastly, I would like to see a table similar to the "extract of the generated dataset" but with 2 more columns. Those columns should show the exact calculated value from the EnergyPlus simulation and the calculated value from the ANN. For rows 0, 1, 2, change Wall U value from 2.256 to 1.112 (everything else same), for rows 3, 4, 5, change Roof U value from 2.865 to 0.869 (everything else same), for rows 6, 7,8, change Orientation from 180 to 45 (everything else same) and from rows 9 and 10, change COP from 2.4 and 3 to 2.7 (everything else the same). As mentioned before, add two more columns and present the actual calculated value from the Energyplus in one column and the calculated ANN value in the other column. The objective is for the reader to double check their calculations against your so that your work can be replicated easily. This table can be presented after Figure 8. in the manuscript.

Response:

The table has been added in the paper. Kindly note S.no. 3 and 8 were taken from the original known validation set. Rest of the cases are from unseen validation data set.

We hope you find the revised manuscript in order submitted for publication in your esteemed journal

Thanking you,

Yours Sincerely
Mannat Kaur

Thank you for addressing my comments. Accepted.