Predicting and Optimizing Building Performance Using Machine Learning

Aleksandra Galczak
Graduate School of Arts and Sciences, Fordham University,
New York City, USA
Ag183@fordham.edu

Abstract: **In the design and optimization of energy-efficient buildings, various characteristics such as surface area, glazing distribution, and orientation play a critical role in influencing performance outcomes, including cooling and heating loads that contribute to building carbon footprint and overall energy consumption. However, accurately predicting these performance metrics remains a challenge due to the complex interplay of multiple features. Traditional methods for evaluating building performance are computationally intensive, time-consuming, and fail to generalize across a diverse range of building designs. Additionally, the lack of actionable insights into the relative importance of design features hinders the ability to make data-driven decisions for optimization. This project aims to address these challenges by leveraging machine learning techniques to develop predictive models that estimate buildings' cooling and heating loads based on key features. By exploring feature importance and relationships, the project seeks to provide valuable insights into optimizing design parameters, thereby promoting sustainable and efficient building designs.**

I. INTRODUCTION

Buildings account for a substantial portion of urban energy consumption, with heating, ventilation, and air conditioning (HVAC) systems being among the most energy-intensive components. As cities like New York implement stricter emissions regulations, including the Climate Mobilization Act and Local Law 97 (LL97)—building owners must adopt more efficient strategies to comply with sustainability mandates. A key challenge in this effort is the early estimation of thermal loads, which traditionally relies on complex and time-consuming calculations using software such as Trane Trace, EnergyPro, and Revit.

This research explores the potential application of machine learning (ML) models in predicting thermal loads before critical design decisions are made—a capability that, if feasible, could transform building efficiency. By leveraging historical data, architectural parameters, and energy simulations, ML models have the potential to forecast thermal demands, allowing stakeholders to make more informed decisions. This approach could enable proactive energy planning, where building owners assess performance before finalizing HVAC designs, leading to optimized operational strategies. Additionally, early predictions could enhance return on investment by guiding more effective financial decisions and reducing long-term energy costs. Architects, too, would benefit from these forecasts, as they could integrate energy-efficient materials that align with projected thermal loads, creating more sustainable designs. However, since this methodology has yet to be fully validated, this paper examines its feasibility by evaluating whether ML-driven load prediction can provide reliable estimates that support energy-conscious building designs while ensuring compliance with LL97. Through a combination of analysis and experimentation, we aim to determine the practicality of ML-based thermal load forecasting in real-world construction and sustainability efforts.

II. OVERVIEW

Modern heating and cooling load calculations rely on sophisticated software like EnergyPro and Trane TRACE, but they still require substantial input from engineers to ensure accuracy. To generate a reliable load estimate, engineers must provide key building-specific and room-specific details. These include location and climate data, which influence outdoor temperature variations, humidity levels, and seasonal heating and cooling needs. Additionally, room orientation plays a significant role in determining how much solar radiation contributes to heat gain, making it essential for engineers to carefully consider each space's exposure to sunlight throughout the day [4].

Beyond external factors, internal heat gains must be accounted for, including occupancy schedules, lighting density, electronic equipment usage, and ventilation requirements. These elements directly affect temperature regulation, as human activity and mechanical systems generate significant heat that must either be retained or dissipated. Building envelope details—such as exterior materials, insulation levels, and thermal resistivity—also play a crucial role in controlling heat transfer and improving overall energy efficiency [4].

One of the greatest challenges engineers face during the early stages of a project is the limited availability of detailed information. In many cases, exact occupancy patterns, material specifications, and HVAC efficiency ratings are not yet defined, forcing engineers to rely on rules of thumb and educated assumptions. These estimates, while useful, can introduce inefficiencies and lead to frequent revisions as more accurate data becomes available. Since this process must be repeated for every room in a building, the overall calculation becomes highly iterative and time intensive [4].

To improve efficiency, emerging technologies are exploring ways to streamline load estimation. AI-driven predictive modeling can refine early-stage assumptions by analyzing historical data and climate trends, reducing reliance on broad approximations. Additionally, integrating information modeling (BIM) software with load calculation tools could automate data transfer, minimizing manual input errors and enhancing precision.

III. DATA

The dataset was generated using Ecotect, an older, discontinued building modeling software. It consists of 768 hypothetical buildings across twelve (12) different building shapes, each varying by eight (8) features, all of them numerical:

It contains architectural and structural parameters that influence thermal loads in buildings. It consists of eight independent variables (X1–X8) that describe physical characteristics of a building and two dependent variables (Y1, Y2) representing its thermal performance.

Independent variables include building compactness, building surface area, wall area, window area, roof area, height, orientation, window area, and window/wall distribution. Dependent variables in the dataset would be heating and cooling loads. See Table A1 in the Appendix for a detailed description of each variable.

This dataset provides a structured foundation to train predictive models, using historical patterns in architectural attributes and their corresponding heating and cooling demands. By leveraging energy simulations and real-world data, ML algorithms could potentially enhance efficiency by anticipating thermal needs early in the design phase—minimizing energy waste, optimizing material use, and ensuring sustainable building performance.

This approach could bring substantial benefits to architects, engineers, and energy planners, enabling more informed decision-making.

IV. DATA PREPROCESSING

The data collected was loaded into Python (Jupyter notebook) and the Sci-kit Learn module [1][2] was used to preprocess data as well as apply machine learning (ML) algorithms. Data collected was simulated data and therefore no missing values or significant outliers were detected. The data distribution is illustrated in figure A1 in the appendix.

A. Feature Selection

Feature selection is crucial to the performance of ML algorithms. However, data used already had a limited number of features and all of them with significant importance, therefore no feature selection was performed.

V. MACHINE LEARNING MODELS

A. Linear Regression

Linear regression was used to model the relationship between dependent variables and independent variables. It assumes a constant rate of change, making it effective for simple trends but limited in capturing complex patterns

To expand the analysis Higher-degree linear regression, or polynomial regression was also applied to the data set and R^2 scores between test and train sets were compared to find the best performing model.

B. Ridge Regression

Ridge regression was used to model the relationship between dependent and independent variables while addressing potential overfitting. Unlike linear regression, which assumes a constant rate of change, ridge regression introduces an L2 regularization term to constrain coefficient magnitude, improving generalization to unseen data. To enhance the analysis, different regularization strengths (alpha values) were applied, and R^2 scores between test and training sets were compared to determine the optimal balance between bias and variance. This ensured that the model maintained predictive accuracy without overfitting to training data.

B. Lasso Regression

Lasso regression was applied to model the relationship between dependent and independent variables while incorporating L1 regularization to enhance feature selection. Unlike linear regression, which allows all variables to contribute freely, lasso regression penalizes large coefficients, forcing some to shrink to zero, effectively selecting the most impactful features. To refine the analysis, different regularization strengths (alpha values) were tested, and R^2 scores for test and training sets were compared to identify the model that achieved the best balance between predictive accuracy and generalization. This ensured optimal performance while reducing unnecessary complexity in the dataset.

C. Neural Network

Neural Network Multi-Layer Perceptron Regressor (MLP) was utilized to model the relationship between dependent and independent variables, leveraging a neural network-based approach to capture complex, nonlinear patterns in the data. Unlike traditional regression models, MLP employs multiple layers of interconnected neurons and activation functions to learn intricate relationships [3]. To optimize performance, various hyperparameters—including L2 regularization alpha were tested. Maximum iterations were set at 2000. Additionally, R-squared scores for test and training sets were compared to ensure the model achieved a balance between predictive accuracy and generalization, preventing issues like overfitting while leveraging the power of deep learning for regression tasks.

D. Performance Metrics

The abovementioned ML algorithms were used to predict cooling and heating loads of a building. For all scenarios, the train-test split was 80/20. Model Evaluation Metrics included R^2 scores for both training and test sets, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE).

VI. RESULTS

A. Linear Regression

When applying linear regression, the first task was to test different polynomial degrees to determine the optimal complexity for accurately predicting thermal loads. The dataset includes multiple independent variables related to building characteristics, such as room orientation, insulation, and occupancy, which are transformed into polynomial features to capture nonlinear relationships. Choosing an appropriate polynomial degree is crucial—too low a degree may lead to underfitting, where the model fails to capture relevant trends in the data, while too high a degree can result in overfitting, meaning the model memorizes patterns in the training data but struggles to generalize to unseen data.

To systematically determine the best polynomial degree, models were trained using degrees 1 through 7, and R^2 scores for both training and test sets were compared. At degree 1, the model assumes a simple linear relationship between independent variables and thermal loads, leading to limited accuracy with an R^2 score of 0.902. This indicates that the model does not sufficiently capture nonlinear dependencies that impact heating and cooling loads. As the polynomial degree increased, performance improved significantly, with degree 4 showing the best balance—training and test R^2 scores remained close together at 0.993 and 0.987, respectively, suggesting strong predictive ability without excessive complexity. However, at degree 5, the gap between training and test R^2 scores widened, signaling overfitting—while the model performed exceptionally well on the training data, it failed to maintain accuracy on unseen data.

To further assess model fit, residual plots were generated for degree 1 and degree 4. Residual plots visualize the difference between actual and predicted values, helping to identify patterns or inconsistencies in the model's predictions. At degree 1, the residuals were unevenly spread and displayed a noticeable trend, indicating systematic prediction errors—suggesting that a simple linear model fails to capture significant variations in thermal loads. In contrast, at degree 4, the residuals were more randomly distributed around zero, demonstrating that the model effectively accounted for nonlinear relationships in the data and minimized systematic errors. This comparison highlights the importance of choosing an optimal polynomial degree to balance predictive performance and generalization.

The comprehensive results for linear regression are listed in Table A2 in the appendix. Figures 3-5 show the best-performing polynomial regression model with degree 4, illustrating its accuracy in predicting heating and cooling loads. Meanwhile, Figures 6-8 present the degree 1 model, showing how a simple linear fit leads to higher residual errors and lower predictive capacity. Additionally, scaling methods, such as Standard Scaler and MinMax Scaler, were applied but did not significantly affect the linear regression model's performance, reinforcing that feature scaling was not a determining factor in polynomial regression accuracy.

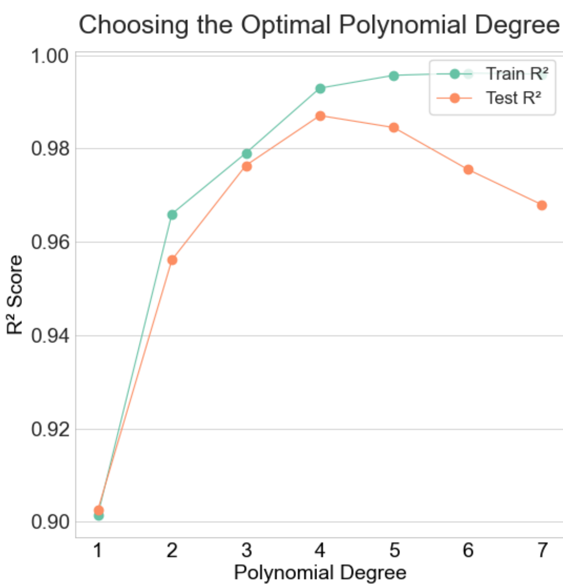

Figure 2. Polynomial degree selection of a Linear Regression.

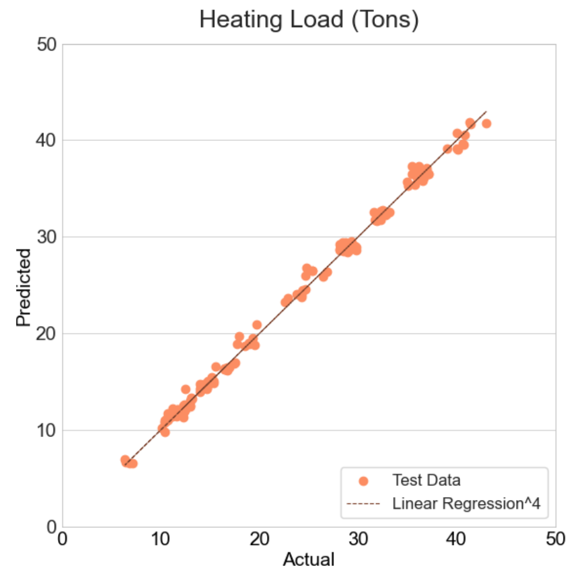

Figure 3. Predicted Heating Load Linear Regression Degree 4.

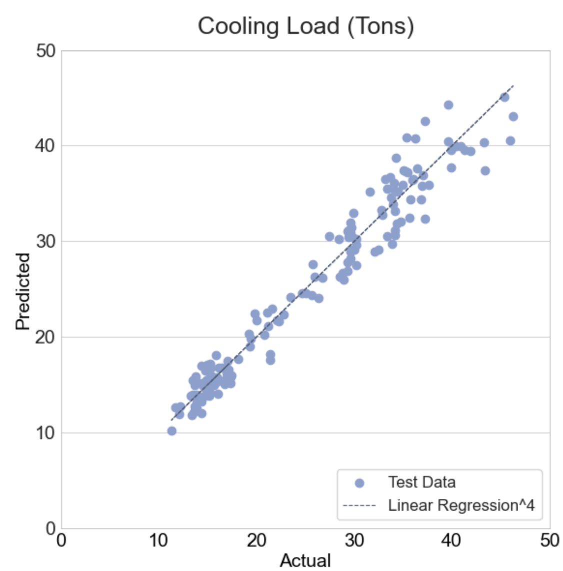

Figure 4. Predicted Cooling Load Linear Regression Degree 4.

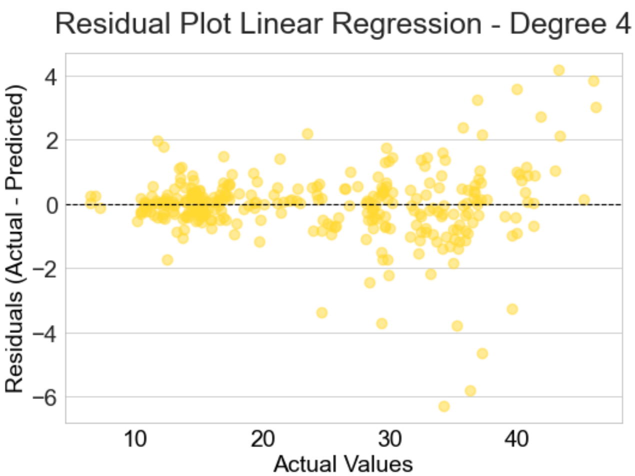

Figure 5. Residual Plot for Linear Regression Degree 4.

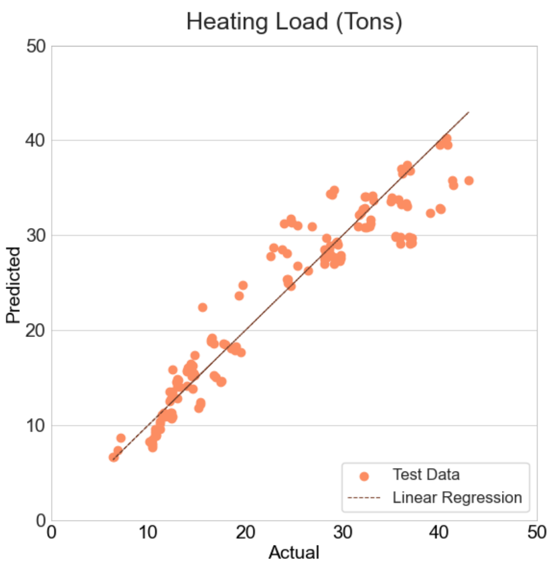

Figure 6. Predicted Heating Load Linear Regression Degree 1.

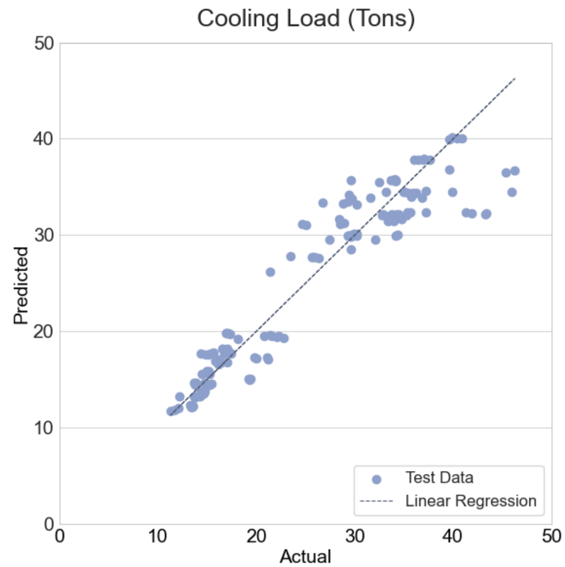

Figure 7. Predicted Cooling Load Linear Regression Degree 1.

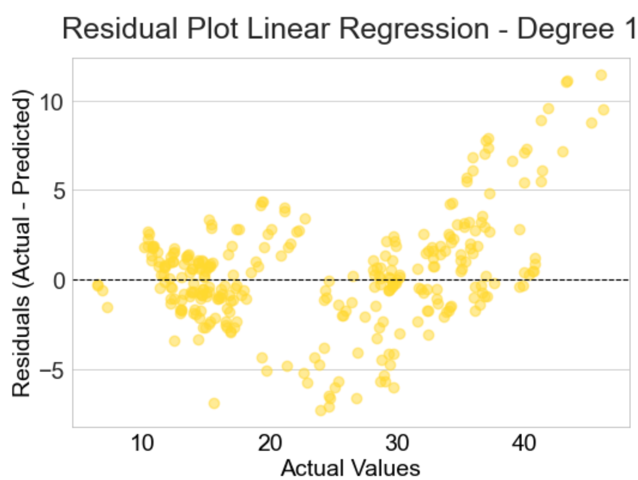

Figure 8. Residual Plot for Linear Regression Degree 1.

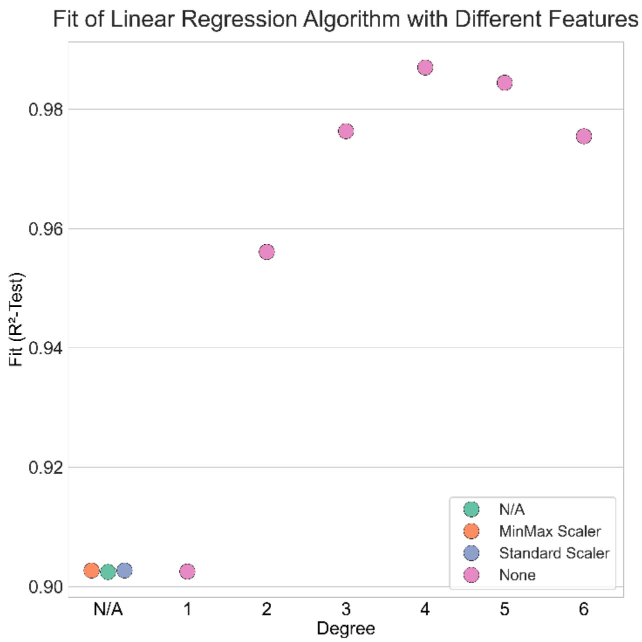

Figure 7. Performance of All Linear Regression Models.

B. Ridge

The Ridge Regression model was tested with various alpha values to evaluate its impact on predictive performance and regularization strength. As alpha increased from 0.01 to 10, both training and test R² scores gradually declined, indicating that stronger regularization reduces model complexity but also slightly lowers accuracy. At alpha = 0.01, the model achieved the highest R² scores (~0.9025 for training and ~0.9014 for testing), suggesting minimal regularization maintains a better fit. However, as alpha increased to 10, the R² scores dropped (~0.8878 for training and ~0.8828 for testing), showing that excessive regularization leads to a simpler, more biased model with reduced predictive power. Additionally, the error values grew as regularization strength increased, reinforcing the trade-off between variance reduction and bias introduction. In summary, lower alpha values (e.g., 0.01) yield better predictive performance, whereas higher values (e.g., 10) promote model generalization at the cost of some accuracy. See Figures 8,9.

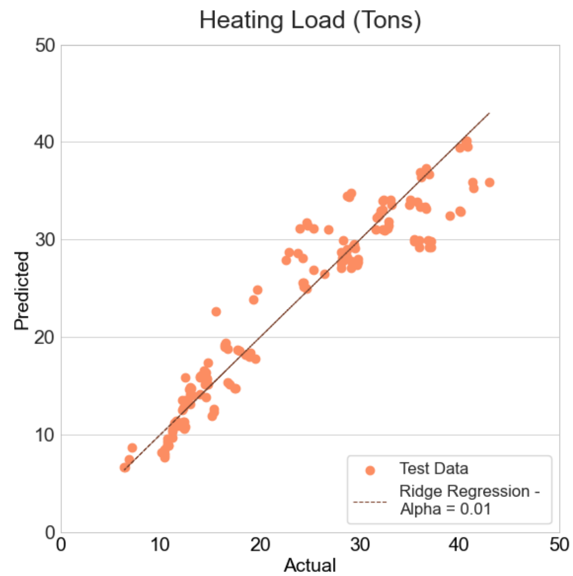

Figure 8. Performance of Ridge Model, Alpha=0.01.

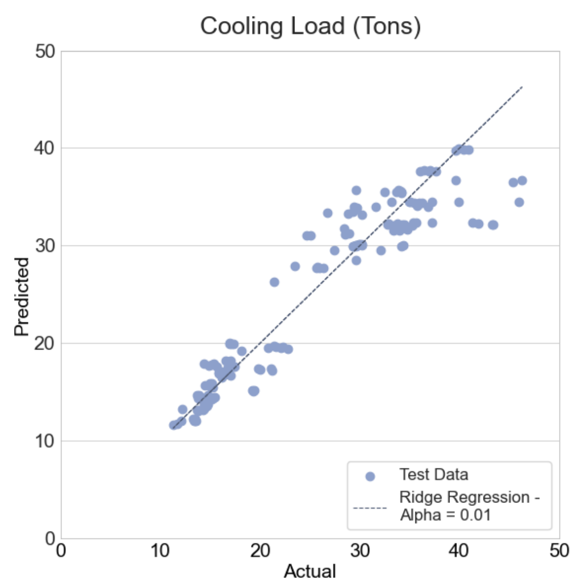

Figure 9. Performance of Ridge Model, Alpha=0.01.

C. Lasso

The Lasso Regression model was tested across different alpha values to evaluate its effect on predictive performance and feature selection. As alpha increased from 0.001 to 10, both training and test R² scores gradually declined, reflecting stronger regularization. At alpha = 0.001, the model achieved the highest R² scores (~0.9025 for training and ~0.9014 for testing), indicating minimal regularization preserved more detailed feature contributions. However, as alpha increased to 1, a noticeable drop in accuracy occurred (~0.7831 for training and ~0.7844 for testing), suggesting that excessive regularization caused feature coefficients to shrink significantly, reducing model complexity. At alpha = 10, the decline persisted (~0.7801 for training and ~0.7819 for testing), further reinforcing the loss of valuable feature information due to aggressive coefficient shrinkage. The error values increased with higher alpha values, highlighting the trade-off between reducing complexity and maintaining predictive accuracy. In summary, smaller alpha values yielded better performance, while larger values led to more constrained models that sacrificed precision for generalization.

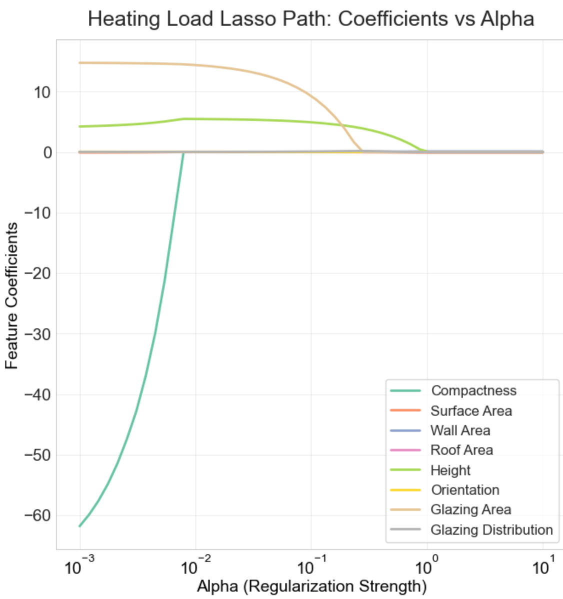

Figure 10. Alpha Regularization for Heating Load.

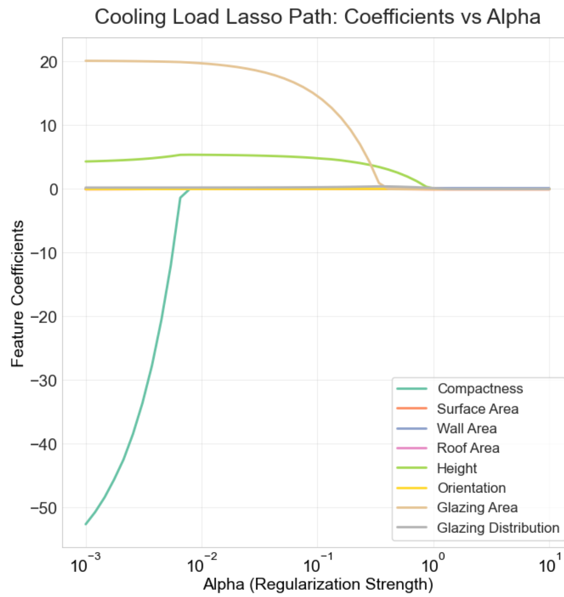

Figure 11. Alpha Regularization for Cooling Load.

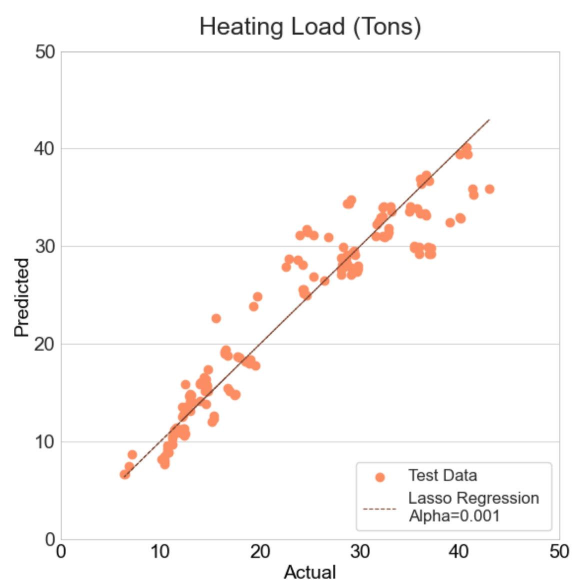

Figure 12. Performance of Lasso Model, Alpha=0.01.

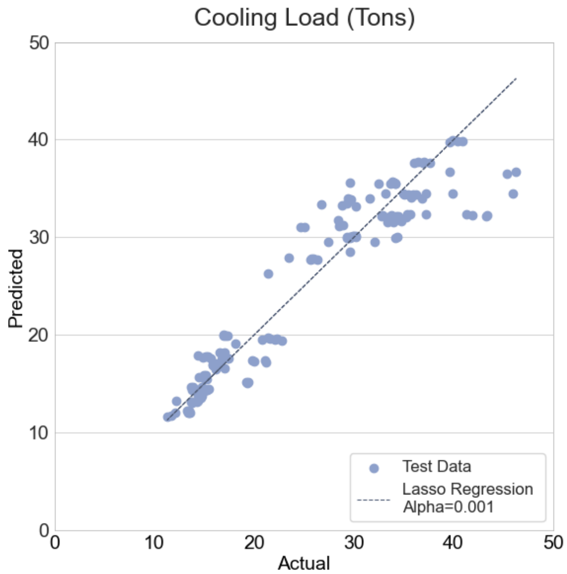

Figure 12. Performance of Lasso Model, Alpha=0.01.

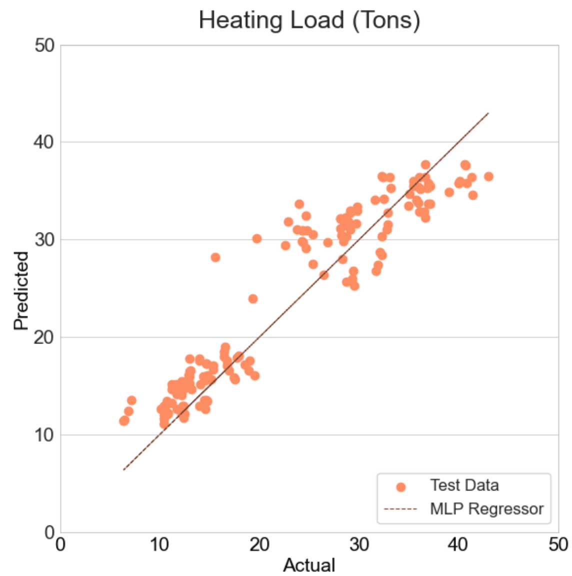

Figure 13. Performance of MLP Model, Alpha=0.01.

D. Neural Network

The Multi-Layer Perceptron (MLP) Regressor was tested across different learning rates (alpha values) to evaluate its performance in modeling heating and cooling loads. As alpha increased from 0.0001 to 0.01, both training and test R^2 scores improved, peaking at alpha = 0.01 with 0.8704 (train) and 0.8618 (test), indicating the model was effectively learning nonlinear relationships. However, when alpha was further increased to 1, R^2 scores dropped to 0.8249 (train) and 0.8098 (test), suggesting excessive learning rate may have negatively impacted performance. Additionally, the error values (MAE and RMSE) followed a similar trend, decreasing initially and then increasing again at alpha = 1, reinforcing that moderate learning rates provided the best balance between accuracy and model stability. In summary, an alpha around 0.01 yielded the best results, while excessively high values introduced inefficiencies, reducing predictive power.

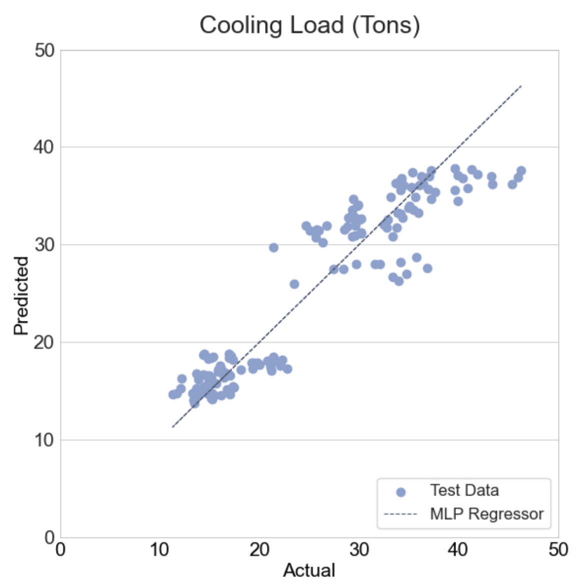

Figure 14. Performance of MLP Model, Alpha=0.01.

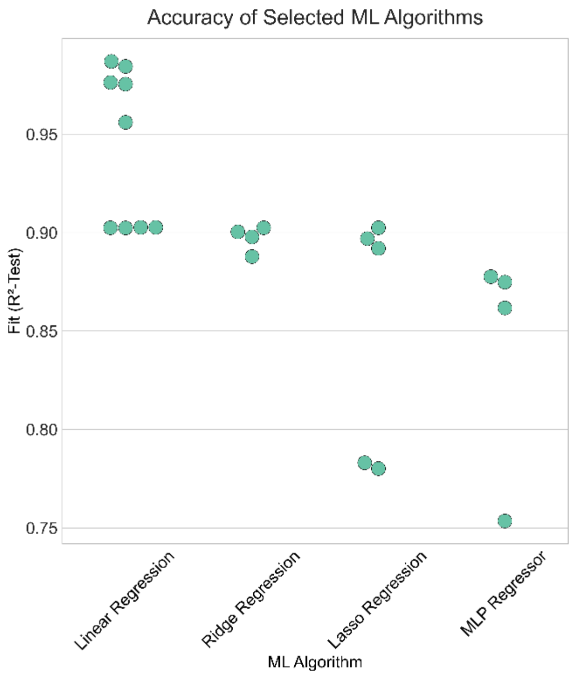

Figure 15. Comparison of different ML Algorithms.

VII. CONCLUSION

This research investigated the feasibility of applying machine learning (ML) models to predict thermal loads before critical design decisions are made, potentially transforming building efficiency and energy planning. By leveraging historical data, architectural parameters, and energy simulations, ML-driven load forecasting offers promising advantages, including proactive energy planning, optimized material selection, and enhanced return on investment.

Through experimentation with Linear Regression, Ridge Regression, Lasso Regression, and Multi-Layer Perceptron (MLP) Regressor, the study evaluated their predictive accuracy using R^2 scores, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) across training and testing datasets. Polynomial regression demonstrated strong accuracy at degree four, though higher degrees led to overfitting. Regularized models such as Ridge and Lasso Regression improved generalization while highlighting the trade-offs between bias and variance. Meanwhile, MLP Regressor exhibited robust performance at moderate learning rates, reinforcing the potential of neural networks for nonlinear load prediction.

However, two key limitations of this study should be noted. First, the dataset did not include thermal resistivity information for any of the building features, preventing architects and designers from making informed comparisons between materials and their effectiveness in improving energy efficiency. Without this data, ML models cannot evaluate how different insulation levels, wall compositions, or glazing types impact overall thermal performance. Additionally, the dataset lacked climate zone classifications, making the predictions location-specific rather than universally adaptable across different geographic regions. Climate plays a crucial role in building energy demand, affecting heating and cooling loads based on factors like humidity, seasonal temperature variations, and solar exposure. Without this consideration, the ML models developed in this study may not generalize well across diverse environments, limiting their applicability in regions with dramatic climate differences.

Overall, results indicate that ML-based thermal load forecasting is feasible, though challenges remain in ensuring reliability for real-world applications. While ML models can generate highly predictive estimates, achieving broad adoption requires improved integration with building simulation tools and real-time sensor data. Future research should focus on refining ML models, incorporating adaptive learning techniques, expanding datasets to include material resistivity and climate zone classifications, and optimizing data collection methods to further enhance accuracy and applicability in architectural and engineering workflows.

REFERENCES

[1] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, Available: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

[2] L. Buitinck *et al.*, “API design for machine learning software: experiences from the scikit-learn project,” *arXiv:1309.0238 [cs]*, Sep. 2013, Available: https://arxiv.org/abs/1309.0238

[3] A.D.Dongare, R.R.Kharde, and Amit D.Kachare, “Introduction to Artificial Neural Network,” 2017. https://www.semanticscholar.org/paper/Introduction-to-Artificial-Neural-Network-A.D.Dongare-R.R.Kharde/04d0b6952a4f0c7203577afc9476c2fcab2cba06

[4] ASHRAE Handbook—Fundamentals, SI ed., ASHRAE, Atlanta, GA, USA, 2021

[5] ASHRAE, "ASHRAE Pocket Guide for Air-Conditioning, Heating, Ventilation, Refrigeration," 10th ed., American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA, 2021.

Appendix A

Variable Name	Role	Type	Description
X1	Feature	Float	Compactness - the ratio of the area of the external partitions to the volume of the building, measures how efficiently a building's shape encloses its volume, impacting energy retention.
X2	Feature	Float	Surface Area - total exterior surface, influencing heat transfer rates.
X3	Feature	Float	Wall Area – the combined area of all walls, affecting insulation properties.
X4	Feature	Float	Roof Area - contributes to heat absorption and dissipation dynamics.
X5	Feature	Float	Height - determines vertical space utilization, influencing natural ventilation.
X6	Feature	Integer	Orientation - The building's directional alignment, affecting solar gain.
X7	Feature	Float	Glazing Area - The proportion of the façade covered by windows, influencing light and heat intake.
X8	Feature	Float	Glazing Distribution Specifies how window surfaces are distributed, impacting thermal balance.
Y1	Target	Float	Heating Load - The energy required to maintain indoor temperatures during cold conditions.
Y2	Target	Float	Cooling Load - The energy needed for cooling in warm conditions.

Table A1. List of features in the clean dataset.

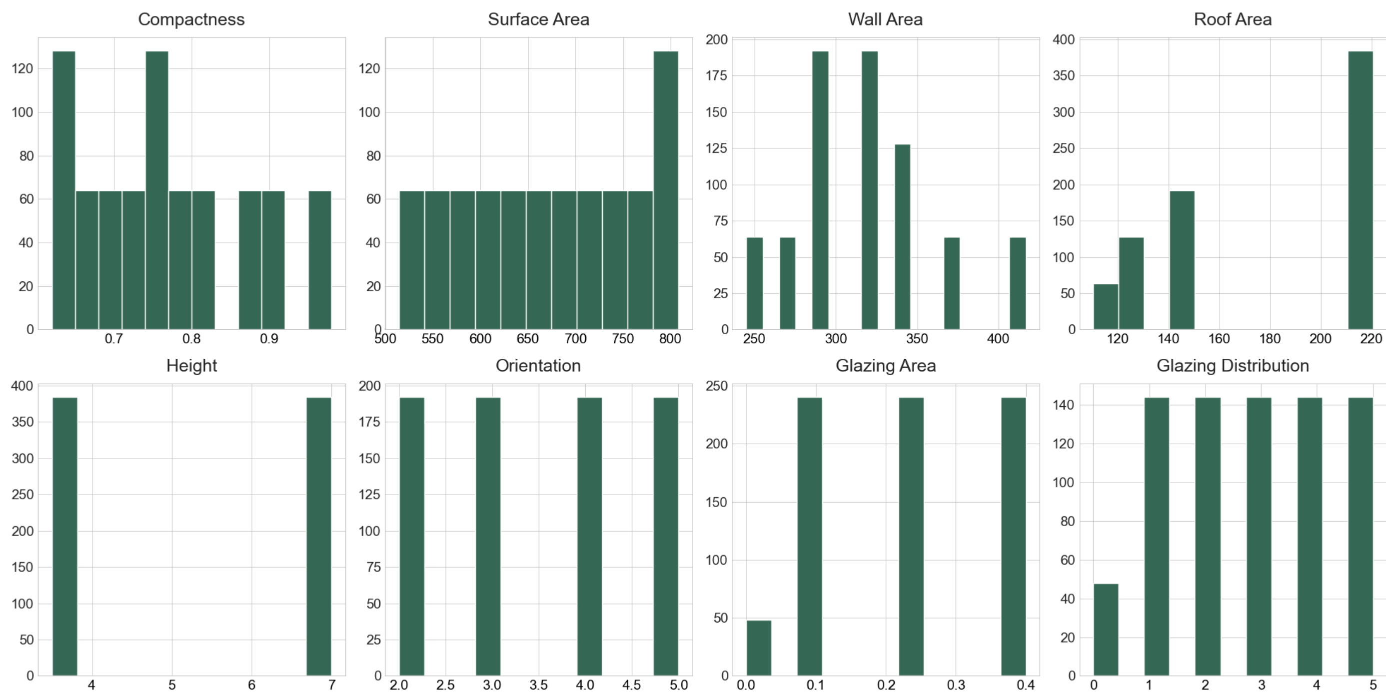

Figure A1. Distributions of features.

Algorithm	Feature Scaling	Degree	Alpha	R^2 - Test	R^2 - Train	MAE	RMSE
Linear Regression	---	---	---	0.9024	0.9010	2.1945	3.0905
Linear Regression	MinMax Scaler	---	---	0.9027	0.9015	2.1895	3.0861
Linear Regression	Standard Scaler	---	---	0.9027	0.9015	2.1887	3.0860
Linear Regression	None	1	---	0.9025	0.9015	2.1922	3.0888
Linear Regression	None	2	---	0.9561	0.9660	1.6283	2.0511
Linear Regression	None	3	---	0.9764	0.9790	1.0194	1.4875
Linear Regression	None	4	---	0.9870	0.9929	0.6789	1.1031
Linear Regression	None	5	---	0.9845	0.9957	0.7245	1.2048
Linear Regression	None	6	---	0.9755	0.9962	0.8216	1.5111
Ridge Regression	None	---	0.01	0.9025	0.9014	2.1913	3.0892
Ridge Regression	None	---	0.1	0.9005	0.8997	2.2358	3.1220
Ridge Regression	None	---	1	0.8979	0.8967	2.3030	3.1622
Ridge Regression	None	---	10	0.8879	0.8828	2.4487	3.3154
Lasso Regression	None	---	0.001	0.9025	0.9014	2.1920	3.0897
Lasso Regression	None	---	0.01	0.8970	0.8964	2.3164	3.1758
Lasso Regression	None	---	0.1	0.8920	0.8899	2.3745	3.2504
Lasso Regression	None	---	1	0.7831	0.7844	3.3495	4.6137
Lasso Regression	None	---	10	0.7801	0.7819	3.3572	4.6467
NN-MLP	None	---	0.0001	0.8152	0.7894	3.4455	4.2612
NN-MLP	None	---	0.001	0.7930	0.7716	3.3986	4.4786
NN-MLP	None	---	0.01	0.7787	0.7298	3.7703	4.6601
NN-MLP	None	---	1	0.8669	0.8582	2.7378	3.6151

Table A2. Results for ML algorithms applied.