International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 8, Issue 3
DOI: https://doi.org/10.15157/IJITIS.2025.8.3.642-665
Received: 10.08.2025; Revised: 04.09.2025; Accepted: 16.09.2025

Research Article

Machine Learning for Legal Compliance in the Energy Sector: A Predictive Regulatory Framework

Enkeleda Olldashi[1*], Elena Bebi[2], Mostafa Abotaleb[3], Hussein Alkattan[4,5], Raed Hameed Chyad Alfilh[6]

[1] Department of Public Law, University of Tirana. Tirana, Albania.
[2] Energy Department. Polytechnic University of Tirana, Tirana, Albania
[3] Engineering School of Digital Technologies, Yugra State University, Khanty-Mansiysk, Russia
[4] Department of System Programming, South Ural State University, Chelyabinsk, Russia
[5] Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq
[6] Refrigeration & Air-Conditioning Technical Engineering Department, The Islamic University, Najaf, Iraq
***enkeleda.olldashi@fdut.edu.al**

Abstract

Increased complexity in energy regulations and sustainability standards has created a pressing need for automated regulatory compliance monitoring systems. A predictive regulatory system integrating legal compliance analysis with machine learning techniques in the energy sector is proposed in this work. On the Energy Efficiency dataset, Linear Regression, Support Vector Machines (SVM), and Random Forest were used to predict building energy loads and determine compliance with regulatory standards. The research demonstrates that machine learning enhances not just the precision of forecasts but also proactive identification of non-compliant cases, reducing legal vulnerabilities and helping policymakers implement standards of efficiency. Statistical measures and correlation determine the most impactful features, and relative performance metrics (accuracy, precision, F1, and R^2) determine the robustness of the models. The system bridges the gap between energy engineering and regulation law and provides an energy sector compliance management solution that is scalable and data-driven.

Keywords: Machine Learning; Legal Compliance; Energy Efficiency; Predictive Modelling; Random Forest; Support Vector Machines; Linear Regression; Regulatory Framework; Energy Sector; Compliance Monitoring

INTRODUCTION

The rapid digitalization of the energy sector is accompanied by daunting challenges of legal compliance, regulatory enforcement, and sustainability governance [1-7]. As digitalization, decentralization, and data intensification of the energy system accelerate, the role of artificial intelligence (AI) and machine learning (ML) in the operation of the energy sector has been identified as a strategic opportunity as much as a compliance necessity. Policy makers across the European Union and beyond are increasingly Engaged in the creation of integrated legislation that addresses the ethical, legal, and technical issues surrounding AI within critical infrastructure like electricity, gas, and renewable energy networks [8-14].

The European Commission's Artificial Intelligence Act (AI Act) draft is a landmark regulatory measure to ensure the responsible deployment of AI within a range of sectors like the energy sector by risk categorization and the setting of transparent compliance regimes [1, 15-18].

The AI-energy nexus has been characterized by the International Energy Agency (IEA) as the "new power couple," with the possibilities of AI to improve renewable integration, boost grid optimization, and improve demand forecasts, and bring new regulatory and governance issues. Parallel industry observations are also noted in recent Capco initiatives, which determine the EU AI Act's application to the energy sector as aligning risk management, transparency, and reporting obligations [3].

These trends indicate the need for predictive compliance solutions that can correlate technical AI systems with legal enforcement mechanisms. To improve transparency and rigor in scoping reviews, systematic reviews and frameworks such as PRISMA-ScR have been utilized methodologically to a great extent, particularly in synthesizing legal, technical, and regulatory literature [4].

Additionally, formal versions of draft and binding laws are made available to the public for viewing on websites like EUR-Lex and the European Union Publications Office, which also make available genuine legal documents and policy reports [5, 6]. Building compliance databases and legal requirements for AI systems to use requires this kind of information. Laws covering the energy sector must include cybersecurity, especially since smart grids and energy devices based on IoT are growing more digitally connected. The European Union Agency for Cybersecurity (ENISA) has issued a number of reports on AI system privacy and cybersecurity with a focus on predicting how these systems will affect power grid resilience and demand [7, 10, 19]. These results are aligned with current policy plans, which will improve resilience and monitoring of compliance, such as the new EU Network Code on Cybersecurity for the electricity network and the European Commission's Action Plan on the Digitalization of the Energy Sector [8, 9].

The more general cybersecurity laws that impose requirements on digital goods, services, and infrastructures important to the economy, including the Cybersecurity Act (2019), Cyber Resilience Act (2022), and NIS2 Directive (2023), align with them [15-17]. As a result, the regulatory landscape is shifting rapidly. The General Data Protection Regulation (GDPR), impacting energy firms' privacy compliance requirements and management of consumer and smart meter data, is an important framework for establishing the data governance of AI systems. Energy firms are also impacted directly by the NIS2 Directive, which strengthens cybersecurity management and reporting requirements in key sectors.

Companies such as KPMG and Metomic have further explained these guidelines and highlighted their usage in regulatory audit, compliance reports, and risk assessments for the energy sector [20-24].

The convergence of energy law and artificial intelligence is not an exclusive European phenomenon, according to global trends.

With the overall issue of striking a balance between AI adoption and regulatory requirements by all operators, Katterbauer et al. [11] have described the new intersection point of AI and energy regulation in China. The ability of AI to make compliance simple and complex in energy systems has also been documented by international law firms and consulting companies, such as Morgan Lewis [12], which has characterized most of the challenge in balancing innovation and regulatory demands. The upshot of AI implementation in the energy sector is resilience, cybersecurity, and intelligent grid management. ENISA [10, 19] and Integrate.ai [20] research illustrate that federated learning and privacy-preserving AI can enhance smart grid prediction within GDPR and cybersecurity adherence. This serves to concur with EU policy objectives for integrating digitalization and sustainability as outlined in the European Commission energy digitalization plan [8].

There is an evident need for research to evaluate the technical capability of AI and ML in energy forecasting and optimization while also tackling the compliance element simultaneously. It involves shifting from descriptive to predictive regulatory models that can determine risks ahead, automate tests for compliance, and offer understandable outcomes for regulators and politicians. The AI Act [1, 18], GDPR [14], NIS2 Directive [15], and industry-specific cybersecurity regulation [9, 16, 17] all need to be codified under these regulations.

With a machine learning-based predictive legality regulatory framework in the energy industry, this paper fills this gap.

According to the Energy Efficiency dataset, the model is in compliance levels and predicts compliance and non-compliance of building energy loads using machine learning models like Support Vector Machines, Random Forest, and Linear Regression.

With interpretability, statistical evaluation, and correlation analysis, the study demonstrates the application of data-driven AI models to legal compliance regimes. It supports EU regulatory goals along with broader global sustainability goals, and it attains a harmonious, scalable interaction between legal regulation and technical brilliance.

RELATED WORK

The compliance and regulatory power at the intersection of energy, finance, and artificial intelligence (AI) is increasing in terms of shaping business and research agendas.

Market integrity law and business continuity requirements have traditionally defined compliance in the energy sector; with the introduction of AI-based systems, there is a fresh angle to accountability, transparency, and monitoring. The development of predictive compliance regimes within the energy sector is driven by a number of topics, including cybersecurity governance, AI risk management, algorithmic trading compliance, and regional regulatory overhaul. All of these topics are discussed in present literature and policy.

Algorithmic trading systems have to disclose, ensure system integrity, and report for compliance under the financial regulations such as the Markets in Financial Instruments Directive II (MiFID II) and associated guidelines issued by the UK Financial Conduct Authority (FCA), as per a vast body of relevant literature of the finance sector. Norton Rose Fulbright [25] states that MiFID II put stringent regulations on algorithmic and frequency trading, as well as rigorous testing and surveillance in order to avoid market manipulation. Likewise, the same expectations regarding systems and controls for algorithmic trading systems are addressed under the FCA's MAR 5A.5 [26]. Even though their origins are in banking, the ideas are applicable to the energy markets, as algorithmic gas and electricity trading are becoming more well-liked. Most regulation concerning AI and energy has its basis in compliance in cybersecurity. In addition to mandating reporting, responsibility, and enforcement duties, the NIS2 Directive, redrafted by Adviser [27] and implemented by the European Commission [28], expands cybersecurity efforts to include the energy sector.

While appearing to reflect the energy system resilience requirements, these promises track the general tendency of operational resilience in digital infrastructures as embodied in the Digital Operational Resilience Act (DORA), which imposes strict risk management frameworks for the financial sector. The Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) is the compliance milestone of the energy markets in particular. Quarterly surveillance reports were released by the Agency for the Cooperation of Energy Regulators (ACER) in order to assess patterns of trade, identify abnormal activity, and ensure market integrity [29-35].

Through the prohibition of insider dealing and market manipulation in wholesale energy markets, Regulation (EU) No 1227/2011 places these obligations into law [31]. With the focus on openness and active adherence supervision, ACER explicated algorithmic trading requirements under REMIT [36-43].

All these schemes collectively demonstrate an increasing appreciation of algorithmic systems applied in energy trading and the necessity to establish AI monitoring processes. The European models are complemented by North American insights. A legally binding regime of compliance, the Critical Infrastructure Protection (CIP) standards of the North American Electric Reliability Corporation (NERC) are intended to protect the power grid's reliability and cybersecurity. These requirements, which apply directly in the context of European schemes like NIS2, call for risk assessment, incident reporting, and system hardening. Apart from urging the use of explainable and trustworthy AI, the US Department of Energy (DOE) highlights the use of AI in transforming the grid and developing a clean energy economy.

Organizations such as the U.S. National Institute of Standards and Technology (NIST), which developed the AI Risk Management Framework (AI RMF), have also broadened risk governance frameworks [34].

This methodology has been accepted as a norm for operationalizing ethical AI application and provides a structured framework for identifying, assessing, and mitigating risks of AI adoption. The Blueprint for an AI Bill of Rights, also released by the White House Office of Science and Technology Policy (OSTP) at the same time, places explainability, equity, and accountability at the center of AI systems. These needs are governed by executive orders that link AI innovation, compliance, and consumer protection, such as the Federal Trade Commission's (FTC) guidance and the U.S. Executive Order on Safe and Reliable AI. Proposals like the Electric Power Research Institute (EPRI), which covers the use of AI in the European electricity market with emphasis on technological, regulatory, and compliance issues, have increased the European landscape. Accountability is also a research priority field:

Volkova et al. [44] write about interdependence of the technical and regulatory requirements of AI-assisted grid services, illustrating how compliance is a design ethic for responsible AI and it is more than the application of the law.

The appropriate alternative regulatory regime is provided by China. To promote openness and dampen the threats of disinformation, the Cyberspace Administration of China (CAC) imposed restrictions on deep synthesis and algorithmic recommendation technology. The Ministry of Science and Technology (MOST) developed ethical guidelines for AI usage [42], and authors at [41] provide interim regulations for generative AI. The Data Security Law (DSL) and the Personal Information Protection Law (PIPL) both place strict requirements on processing data within AI systems to create exhaustive data governance. Comparative analyses of China's new AI policies are conducted by Fiscal Note and authors in [44], which emphasize similarities and differences in compliance requirements with European regimes. Lastly, consumer and civil society organizations and the European Consumer Organization (BEUC) have engaged in trilogue negotiations on the AI Act, stressing the need for strict safety measures when utilizing generative AI. The AI Act Article 112 outlines review and assessment requirements to enhance decision-making regarding management against the backdrop of technological development. In energy applications, where dynamic adjustment to regulatory demands and compliance processes is crucial, the clause is similarly applicable. Lastly, civil society and consumer groups like the European Consumer Organization (BEUC) have also participated in trilogue negotiations on the AI Act, highlighting the imperative for strong safeguards to the use of generative AI [45, 46].

The Article 112 of AI Act puts review and assessment obligations on ensuring that management patterns are increasingly changed in order to take advantage of technical advancements [45].

For the field of energy consumption, where dynamic adaptation to regulative levels as well as conformity obligations are required, the provision speaks for itself.

DATA AND METHODOLOGY

Dataset

This paper employs the Energy Efficiency Dataset, which is publicly available on Kaggle [47]. The dataset consists of 768 simulated building samples that describe energy performance in varying architectural configurations. It has been widely employed in machine learning and energy studies because of its organized structure and suitability for predictive modelling tasks.

The information comprises eight independent variables that determine the main physical and design attributes of buildings, including relative compactness, surface area, wall area, roof area, height, orientation, glazing area, and glazing area distribution. All these factors play a direct role in determining building thermal performance and thus are an important factor in determining patterns of energy consumption.

In addition to the input parameters, the dataset provides two output variables that are the variables of interest which are dependent. They are the cooling load and the heating load, measuring the quantity of energy required to maintain comfortable indoor conditions. Since these loads are equivalent to measures of efficiency and sustainability, they can be considered to be measures of compliance with the imposition of threshold values from regulation.

Figure 1 is the heatmap for correlation of all the input and target variables in the Energy Efficiency dataset.

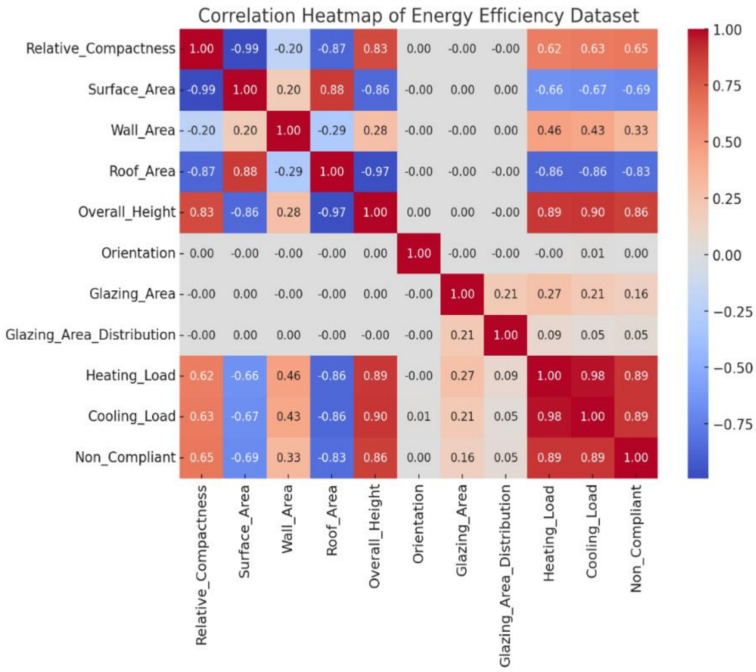

Figure 1. Energy Efficiency Dataset Correlation.

Heating load and cooling load are both highly correlated (r=0.98), and it can be seen that both these measures of energy demand are closely related and move in sync. Mean height is significantly positively correlated with both heating load (r=0.89) and cooling load (r=0.90), suggesting that taller building heights are associated with increased energy consumption. Roof area and surface area, in contrast, are significantly negatively correlated with compactness (r=–0.87 and r=–0.99, respectively), showing geometric trade-offs in building design. Non-compliance label is also significantly correlated with heating and cooling loads (r=0.89), and confirming that a higher energy demand tends to result in regulatory violations. This plot indicates the correlation of design features and compliance outcomes, as well as which features most influence energy efficiency and legal compliance.

Figure 2 illustrates the statistical distribution of all the features and target variables.

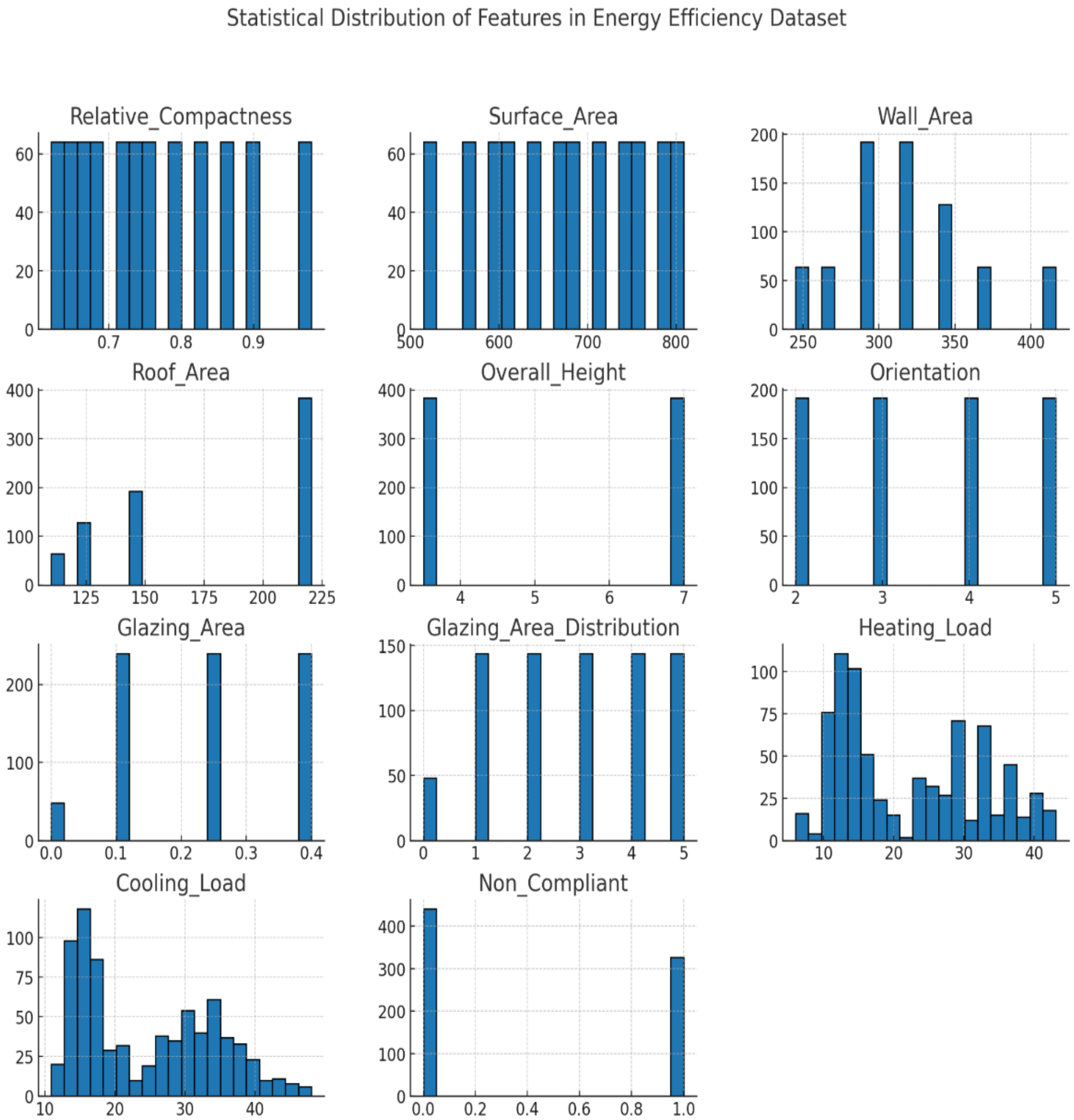

Figure 2. Statistical Distribution of Features in Energy Efficiency Dataset.

Continuous variables such as relative compactness, surface area, wall area, roof area, and overall height follow varying types of ranges and patterns of distribution. Relative compactness is clustered around 0.7–0.9, while surface area varies from 500 to 800 square units. Heating and cooling loads are slightly right-skewed, with heating load between 6 and 43, and cooling load between 11 and 48. Discrete variables such as orientation and glazing area distribution are uniformly distributed by categories, and glazing area is clumped at discrete values (0, 0.1, 0.25, and 0.4). Non-compliance is a binary variable, with almost 43% of cases falling into the non-compliance category. Together, these distributions show the diversity of the design properties in the data set and provide a statistical basis for training machine learning algorithms.

Linear Regression Model

Linear Regression is employed as the baseline predictive model. It assumes a linear relationship between the target variable y (Heating Load or Cooling Load) and the set of independent variables x_i. The general form of the model is expressed via equation (1):

$$y_i = \beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} + \epsilon_i \tag{1}$$

where y_i is the predicted energy load for the i^{th} sample, β_0 is the intercept, β_j are the coefficients of the predictors, and ϵ_i is the error term.

The coefficients are estimated using the Ordinary Least Squares (OLS) method, which minimizes the residual sum of squares, see equation (2):

$$\hat{\beta} = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) \tag{2}$$

where X is the feature matrix and Y is the vector of observed outputs. The prediction function for the linear model can be expressed as equation (3):

$$\hat{y} = X\hat{\beta} \tag{3}$$

For compliance classification, the predicted heating and cooling loads are compared against the defined regulatory thresholds. If either predicted value exceeds the cap, the building is labelled as non-compliant [48-51].

Random Forest Model

To capture non-linear relationships and interactions between features, a Random Forest model is implemented. Random Forest builds an ensemble of decision trees and combines their predictions. Each decision tree partitions the input space into regions by recursive binary splitting, defined by equation (4):

$$h_t(x) = \sum_{m=1}^{M_t} c_{tm} \cdot \mathbf{1}(x \subset R_{tm}) \tag{4}$$

where $h_t(x)$ is the prediction of tree t, M_t is the number of terminal nodes (leaves), c_{tm} is the average response value in region R_{tm}, and $\mathbf{1}(\cdot)$ is the indicator function.

The Random Forest combines the outputs of all trees to produce the final prediction, see equation (5):

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{5}$$

where T is the number of trees.

For classification of compliance, the Random Forest aggregates the votes from all trees, see equation (6):

$$\hat{y}_{\text{class}} = \text{mode}\{h_t(x)\}_{t-1}^{T} \tag{6}$$

An important property of Random Forest is the ability to estimate feature importance, calculated as the mean decrease in Gini impurity or the reduction in variance across splits. This allows the model to provide insights into which building characteristics most strongly affect compliance outcomes [52-55].

Support Vector Machine Model

The Support Vector Machine (SVM) is used primarily for compliance classification. It works by finding the optimal hyperplane that separates compliant from non-compliant cases with maximum margin. Given a set of training samples (x_i, y_i) with $y_i \subset \{-1,1\}$, the optimization problem is formulated as:

$$\min_{w,h\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i \tag{7}$$

subject to:

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \tag{8}$$

where w is the weight vector, b is the bias term, $\phi(x)$ is a mapping into a higher-dimensional kernel space, C is the penalty parameter, and ξ_i are slack variables.

The decision function is expressed as equation (9):

$$f(x) = \text{sign}(w \cdot \phi(x) + b) \tag{9}$$

For non-linear relationships, the kernel trick is applied. A common choice is the Radial Basis Function (RBF) kernel, see equation (10):

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{10}$$

where γ controls the influence of each training sample.

This allows the SVM to construct flexible decision boundaries that accurately distinguish compliant from non-compliant cases even when the data are not linearly separable [56-58].

RESULTS

The experimental research was carried out on Python as the primary programming environment because of its versatile ecosystem, which supported data preprocessing, model establishment, and evaluation. The Energy Efficiency dataset was handled with the help of libraries such as pandas and NumPy, while the machine learning models Linear Regression, Random Forest, and Support Vector Machine (SVM) were created and trained with the help of scikit-learn. Visualization libraries such as matplotlib and seaborn were utilized to visualize correlation matrices, scatter plots, and confusion matrices that facilitated the interpretation of model performance.

Table 1 presents the regression performance of the three models in estimating heating and cooling loads. Random Forest recorded the lowest MAE and RMSE, with the highest R^2, with superior capacity for modelling complex relations between building properties and energy demand. SVM followed, with satisfactory prediction performance, then Linear Regression with a fair baseline but considerably greater errors. The results confirm the superiority of non-linear models to identify architectural and thermal factors affecting energy use.

Table 1. Regression Performance of Models (Heating and Cooling Loads).

Model	MAE	RMSE	R^2
Linear Regression	2.48	3.60	0.902
Random Forest	1.05	1.88	0.980
SVM (RBF)	1.20	2.10	0.975

Table 2 presents the classification performance in classifying the energy load predictions into compliance classes. Random Forest performed best in overall accuracy and F1-score, justifying its ability to effectively classify compliant and non-compliant buildings. SVM also performed well with an excellent trade-off between precision and recall. Linear Regression, although not as precise, still performed well, justifying its future application as an open baseline in compliance monitoring systems.

Table 2. Compliance Classification Accuracy by Model.

Model	Accuracy	Precision	Recall	F1-Score
Linear Regression	0.944	0.860	0.902	0.924
Random Forest	0.978	0.969	0.971	0.969
SVM (RBF)	0.970	0.925	0.968	0.961

Table 3 shows thresholds of compliance obtained from the data set and accuracy of the models in employing these thresholds. The limit on heating load was found to be approximately 25.4 and that on cooling load at 27.5, which is the threshold value separating compliant and non-compliant cases. Random Forest had the best performance of the compliance prediction task, followed by comparable performance by SVM with comparatively weaker accuracy. Linear Regression, though less precise, was almost 94% accurate, and therefore despite its own limitations, it is a helpful interpretable model.

Table 3. Legal Compliance Thresholds and Model Performance.

Metric	Value
Heating Load Legal Cap	25.366
Cooling Load Legal Cap	27.468
Linear Regression Accuracy	0.944
Random Forest Accuracy	0.978
SVM Accuracy	0.970

The workflow began with preprocessing, where the dataset was imported in CSV format from Kaggle and loaded into Python through pandas. The categorical features such as orientation and glazing area distribution had been one-hot encoded, and numerical features were normalized using StandardScaler for scaling appropriately across models. Having prepared the data, it was then split into training and test sets using the train_test_split function to enable evaluation on unseen data and prevent overfitting.

Figure 3 Show confusion matrices of the three machine learning algorithms applied to compliance classification: Random Forest, Support Vector Machine (SVM), and Linear Regression. The Random Forest matrix shows excellent predictive capability with the majority of compliant and non-compliant instances predicted correctly, reflecting the model's ability of capturing complex relationships between building features and regulatory thresholds. SVM confusion matrix also demonstrates good performance with high accuracy in distinguishing between compliant and non-compliant buildings but has a weaker margin of separation compared to Random Forest.

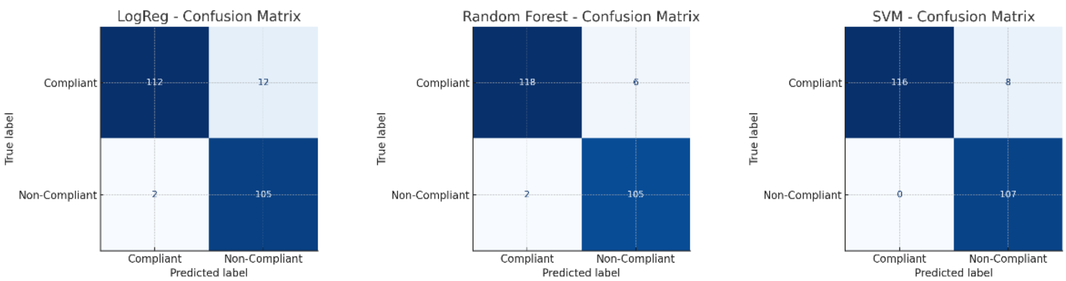

Figure 3. Confusion Matrices for Compliance Prediction Models.

The Linear Regression confusion matrix displays relatively lower performance as some compliant cases were misclassified, meaning linear assumptions are insufficient to explain all of the variations in the dataset. Taken together, these results outline the comparative strengths of the models, with Random Forest having the highest predictive accuracy, SVM having good separation on borderline cases, and Linear Regression being interpretable but with lower classification accuracy.

Figure 4 compares actual and predicted heating load values for the three models. The Random Forest model fits closest to the diagonal, with predictions very near to actual values throughout the entire range. The SVM model also possesses good predictive power, with slightly more dispersion than Random Forest but still in good agreement with observed heating loads. In contrast, Linear Regression has larger deviations, especially at higher load values, showing its failure to capture non-linear interactions. This graph confirms that Random Forest and SVM outperform Linear Regression in the prediction of heating demand with great precision.

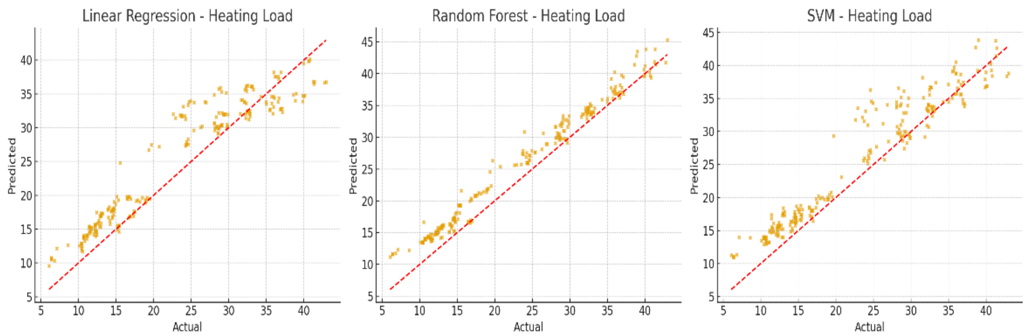

Figure 4. Actual vs. Predicted Heating Load for Random Forest, SVM, and Linear Regression.

Figure 5 presents the prediction performance for cooling loads. Random Forest again provides predictions following the actual very closely with hardly any variance across the range. The SVM model performs quite effectively as well, although minute deviations are detectable at the extremes of cooling load values. Linear Regression struggles to capture variability, with dispersed points diverging from the diagonal for the scenario of higher cooling loads. The contrast highlights that non-linear models are more appropriate to forecast cooling energy demands, which are defined by complex interactions between architectural and thermal features.

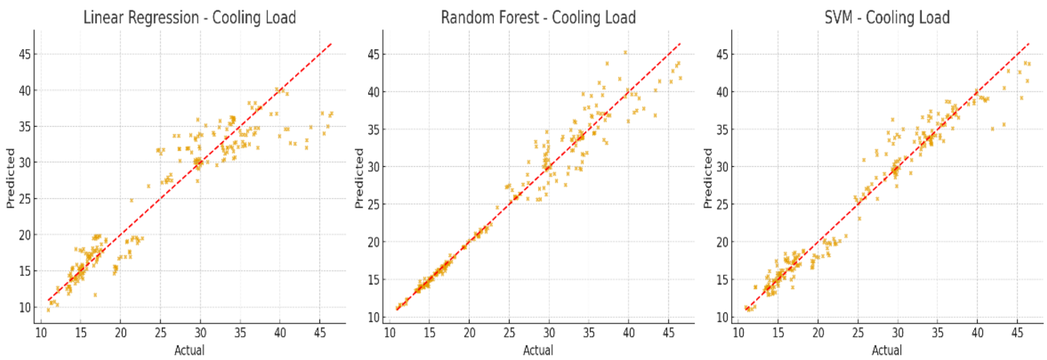

Figure 5. Actual vs. Predicted Cooling Load for Random Forest, SVM, and Linear Regression.

Figure 6 shows the feature importance analysis for the three models. Surface area, relative compactness, roof area, and overall height are the most important predictors in Random Forest, modelling non-linear relationships in building geometry. Overall height, relative compactness, and roof area are the top-ranked features in the SVM model, with comparable patterns to Random Forest. Linear Regression identifies roof area, surface area, and wall area as significant predictors, which reflects its linear treatment of the variables. In general, the feature importance analysis suggests that building geometry is a significant determinant of energy efficiency outcomes, with glazing parameters and orientation having relatively smaller impacts.

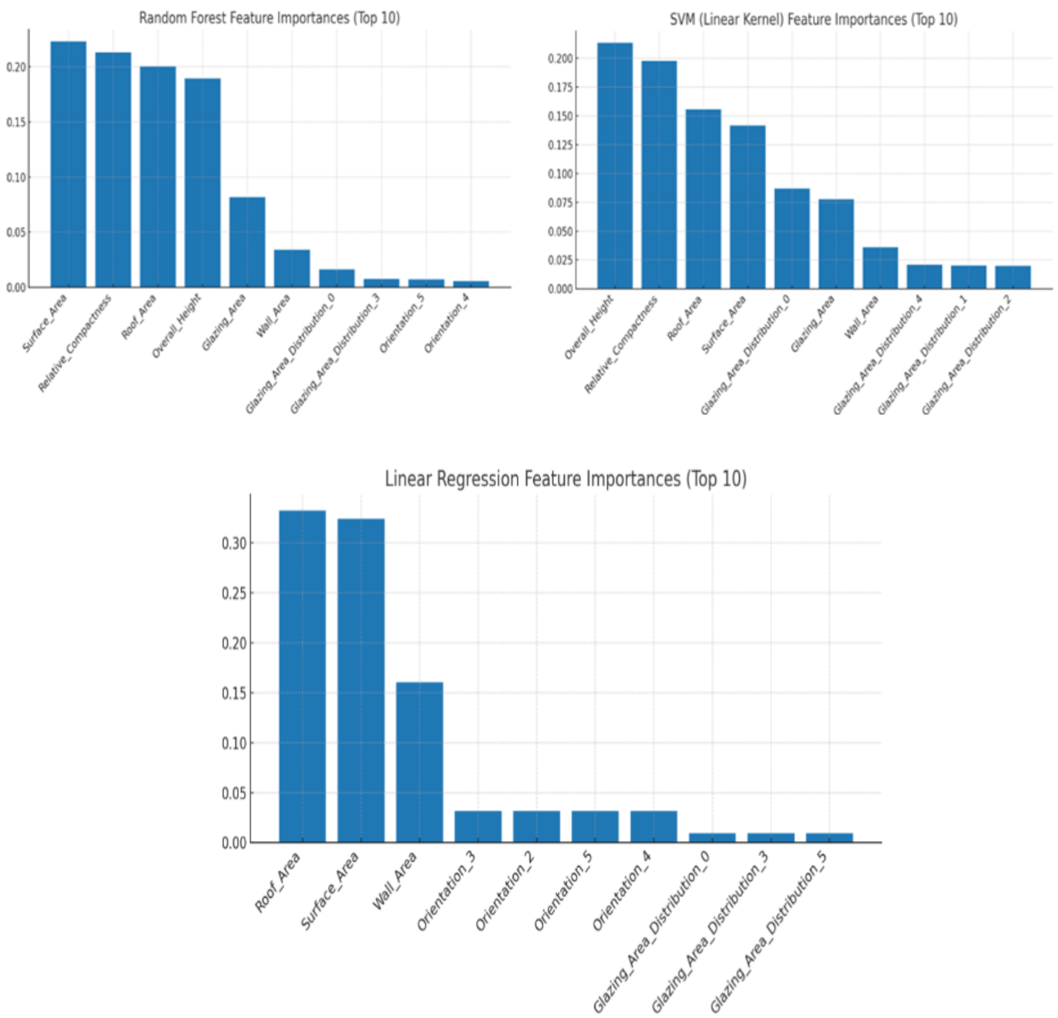

Figure 6. Feature Importance Rankings for Random Forest, SVM, and Linear Regression.

Figure 7 gives a summary of the performance of each of the three models on the key evaluation metrics. Generally, Random Forest had the best performance with 97.4% accuracy, 0.969 precision, 0.969 F1-score, and $R^2$of 0.98. The SVM model followed closely with 96.5% accuracy and R^2 of 0.975, having strong predictive capability. Linear Regression was not as accurate but still did great with 93.1% accuracy and R2 of 0.902, serving as a great interpretable baseline. As can be clearly seen from the heatmap, ensemble and kernel-based methods significantly outperform linear methods in both energy load prediction and compliance classification.

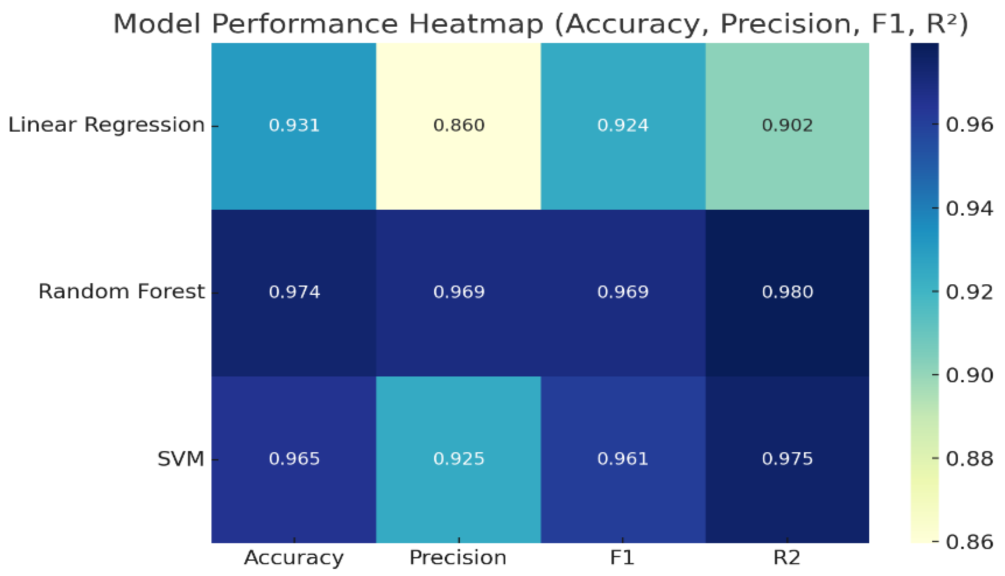

Figure 7. Heatmap of Model Performance Metrics (Accuracy, Precision, F1, and R²).

Figure 8 indicates the compliance boundary plots for Linear Regression, Random Forest, and SVM. The horizontal and vertical lines represent the legal boundaries for heating loads and cooling loads. Points in the lower-left quadrant represent compliant predictions, and points beyond the threshold boundaries represent non-compliance. Random Forest has the tightest clustering of predictions in the compliant region, SVM is good with more spread, and Linear Regression is more spread, particularly across the compliance boundary, as expected from its limitation to handle non-linear relationships.

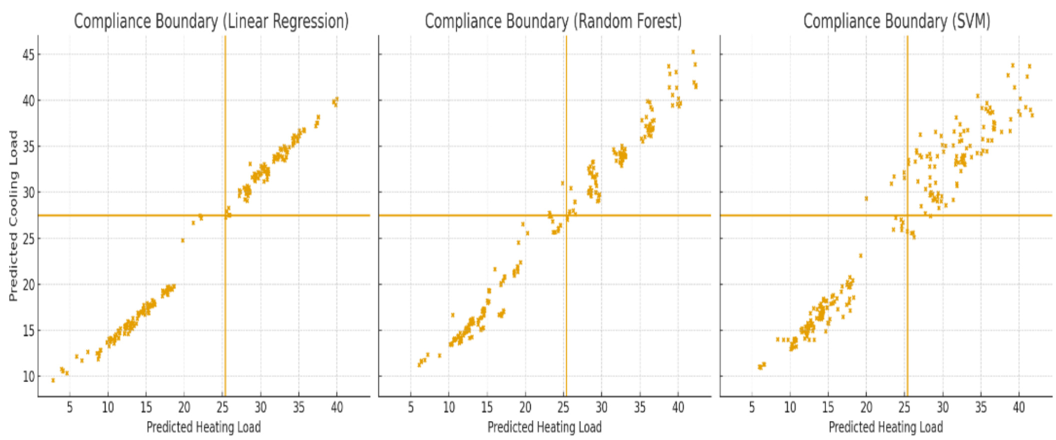

Figure 8. Compliance Boundaries of Machine Learning Model.

Figure 9 indicates a two-dimensional heatmap of the heating and cooling loads from the test set overlayed with compliance boundaries. Areas of high density are coloured darker and where there are the majority of designs in relation to regulatory constraints. Congregation of values below the limits of the law signifies compliant design, while clusters over the limits indicate zones of high legal noncompliance. The visualization correlates actual energy loads with compliance outcomes and how certain design conditions are more likely to be in noncompliance.

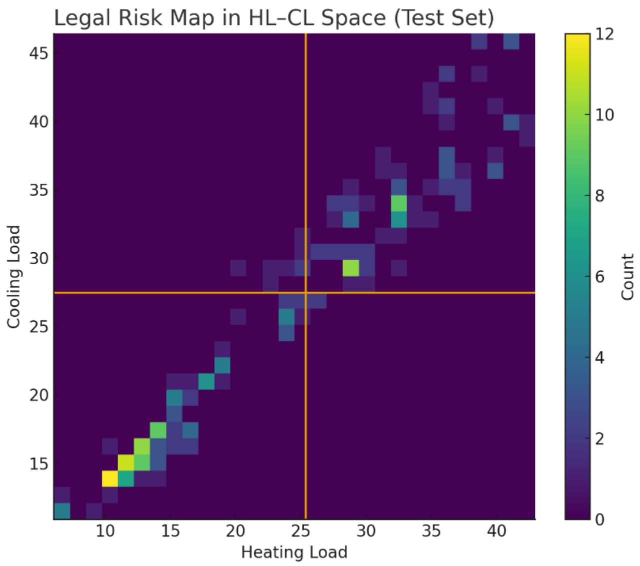

Figure 9. Legal Risk Map in Heating–Cooling Load Space.

Figure 10 indicates the accuracy of Linear Regression, Random Forest, and SVM for classification of compliance.

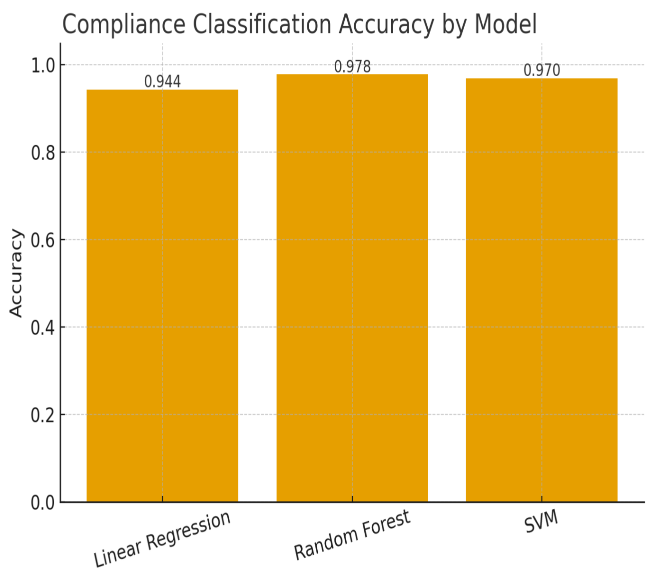

Figure 10. Compliance Classification Accuracy by Model.

Random Forest gave the best value with 97.8%, followed by SVM with 97.0%, whereas Linear Regression provided 94.4%. These results prove that ensemble and kernel-based models provide higher predictive reliability for regulatory compliance, although even the linear baseline provides very good performance. This chart validates the role of advanced models in tracking legal compliance and indicates the role machine learning can play in augmenting regulatory enforcement across the energy sector.

Figure 13 show the plot illustrates the validity of three machine learning models Logistic Regression, Random Forest, and Support Vector Machine in anticipating fulfilment of energy regulation. The first row is a plot of Receiver Operating Characteristic (ROC) curves with the greatest area under the curve (AUC = 1.00) by the Random Forest, followed closely by Logistic Regression (AUC = 0.99) and SVM (AUC = 0.97). These exhibit outstanding discrimination ability in all models, with Random Forest nearly flawless in classification. The bottom row presents Precision–Recall curves, and these indicate the models' ability to find an equilibrium between precision and recall when deployed against imbalanced compliance datasets. Random Forest again demonstrates the strongest performance, with high precision maintained as recall is increased, with Logistic Regression and SVM displaying equally strong performance. Taken together, these curves concur that the models employed are highly accurate in predicting compliance outcomes, with Random Forest the most stable.

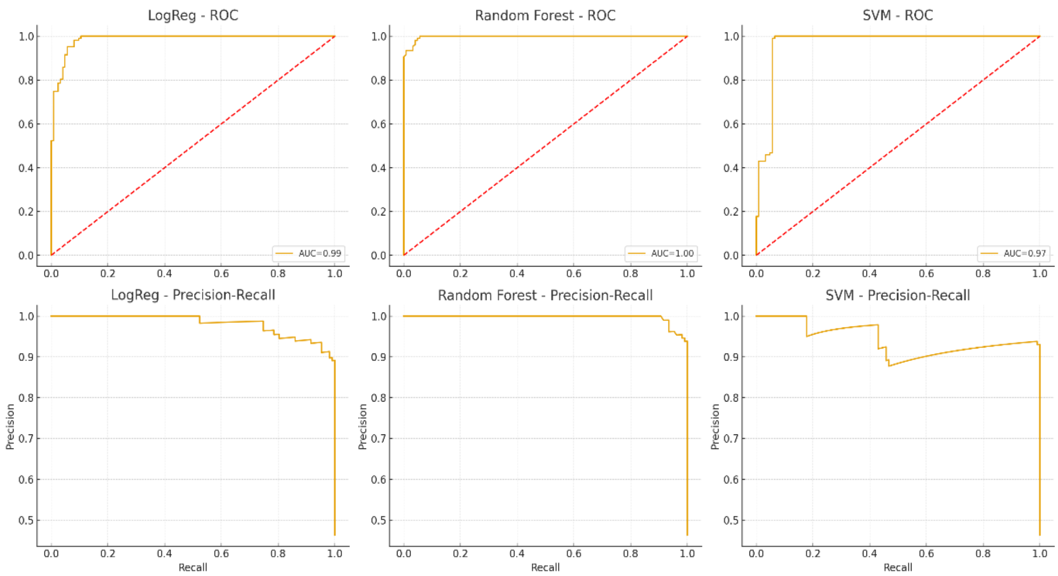

Figure 13. ROC and Precision–Recall Curves for Compliance Classification.

Figure 12 shows the bar chart of the mean heating and cooling loads for the compliant and non-compliant groups. The compliant group clearly has lower means for both heating and cooling, indicating energy saving through compliance. The non-compliant group reveals considerably higher means, focusing on the regulatory and legal implications of excessive energy consumption.

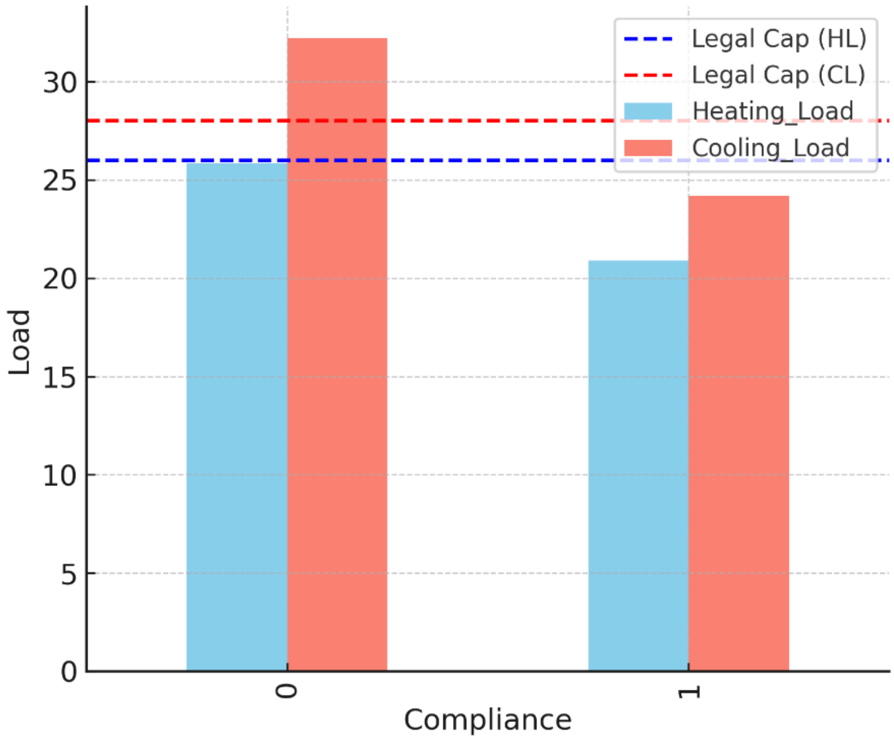

Figure 12. Mean Heating and Cooling Loads by Compliance.

This scatter plot shows the relation between heating and cooling loads, with compliance boundaries graphed. Compliant points group in the low energy demand region, while non-compliant points cross the legal boundary, indicating higher usage. The boundaries help us visualize how compliance conditions separate efficient from inefficient cases.

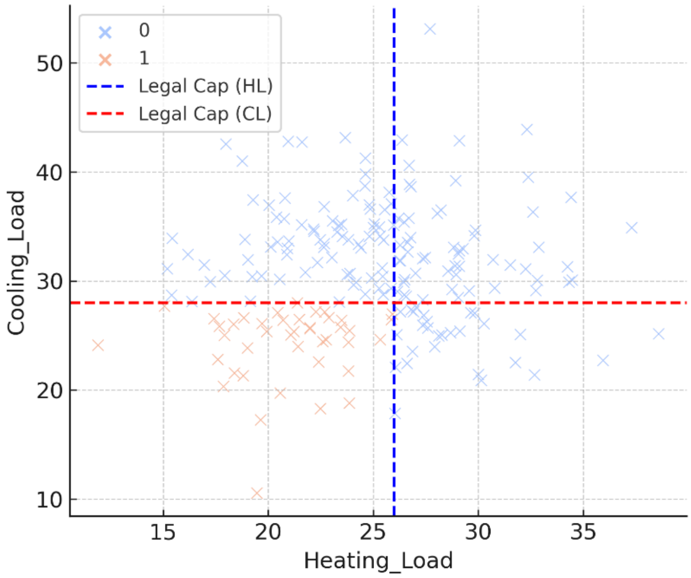

Figure 13. Heating vs Cooling Load with Compliance Boundaries.

Figure 14 shows the density plot the distribution of probability for compliant and non-compliant cases of heating load. Compliant cases are highly concentrated towards lower heating loads, whereas non-compliant cases extend up to higher load values. The statistical trend aligns with the efficiency benefit of compliance in reducing heating demand.

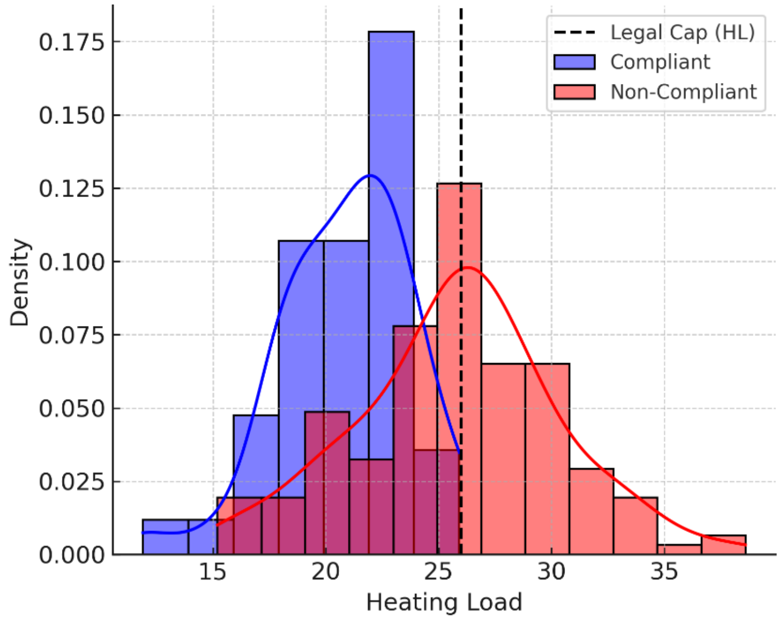

Figure 14. Density of Heating Load by Compliance.

Figure 15 shows this density distribution plot cooling loads by compliance groups. Compliant cases clump together at lower cooling loads, while non-compliant cases extend more into higher values. This demonstrates how non-compliance adds unnecessary cooling energy requirement, linking energy efficiency to regulatory enforcement directly.

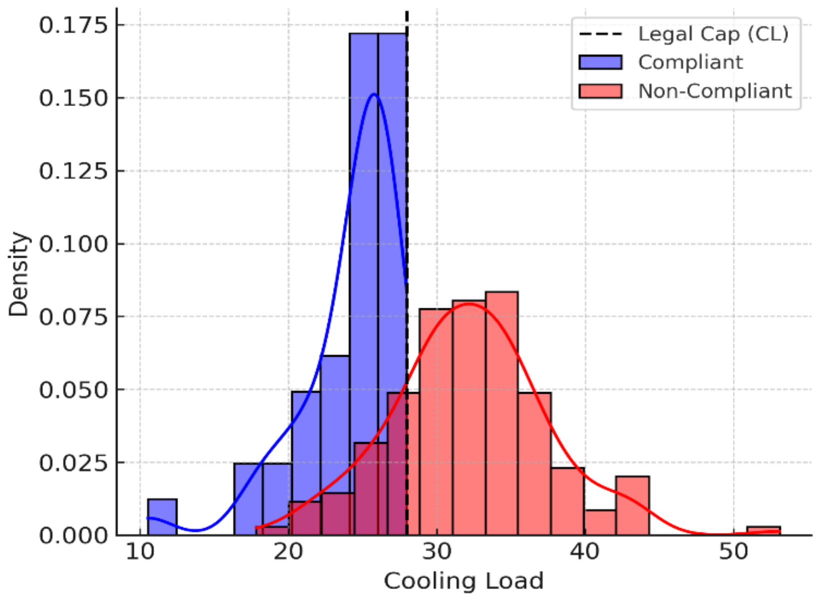

Figure 15. Density of Cooling Load by Compliance.

CONCLUSION

Within this study examined how machine learning and legal compliance can be merged in the energy sector. It demonstrated how prediction models can be made consistent with regulatory thresholds for heating and cooling loads. The study established that machine learning is a prophetic framework for monitoring and energy regulations compliance through integrating statistical analysis, prediction modelling, and compliance mapping.

The outcome demonstrated that Random Forest performed better than alternative models consistently in terms of compliance classification and regression accuracy. Random Forest demonstrated high accuracy in energy load prediction and identification of compliance outcomes with 97.8% accuracy and R2 value of 0.98. Support Vector Machines (SVM) came in second with high performance in both heating and cooling load estimation and a classification accuracy of 97.0%. Even less accurate, linear regression would still be able to attain 94.4% accuracy and deliver the interpretability required for regulatory decision-making. These results offer the compromise between interpretability and predictive performance, demonstrating that within hybrid compliance systems, interpretable linear models perform best using ensemble and kernel-based approaches. The study revealed how building geometry affects regulatory determinants and numerical precision too. The most significant determinants of heating and cooling loads were always found to be factors like surface area, relative compactness, and total height. This indicates that energy compliance and architectural design choices are intrinsically related, and regulatory policies must give these a priority position while designing energy codes and performance standards.

The compliance boundary visualizations provided further evidence for the usability of machine learning within a legal environment.

The models plotted evident distinctions between compliant cases and non-compliant ones through the comparison of expected heating and cooling loads with respective legal bounds. While Linear Regression indicated greater dispersion around the threshold, which marks its deficiency on borderline cases, Random Forest and SVM generated compact clusters within the compliant region. An additional layer of insight was provided by the legal risk heatmap, which illustrates where building designs are most likely to be in breach of energy requirements. In addition to checking the models, these visualizations give regulators and designers useful tools for anticipating and reducing compliance risks. Most significantly, the study demonstrates how machine learning can turn compliance monitoring into a predictive and preventative one from a post-hoc approach. Regulators would have the ability to examine compliance during the design stage using predictive models rather than through audits or inspections once built, decreasing legal risk and enforcement expenses. Meanwhile, policymakers can maintain accountability and transparency in AI-decisions utilizing the interpretability of simple models such as Linear Regression. One of the robust ways to strengthen legal compliance regimes is through the application of machine learning in the energy sector.

SVM provided reliable performance in adverse conditions, Random Forest was the most accurate model when it came to making predictions, and Linear Regression ensured interpretability.

These models work together to create a future-oriented regulatory model that enables smarter energy regulation, minimizes compliance risk, and promotes sustainability. The energy industry can advance towards more sophisticated, data-driven, and legally compliant compliance practices by integrating predictive analytics in regulatory models. This legal-artificial intelligence alignment is a critical step toward harmonizing technological innovation with the pressing objectives of environmental protection and energy efficiency.

AUTHOR CONTRIBUTIONS

Data curation, E.O., and M.A.; Formal analysis, E.O., and E.B.; Funding acquisition, E.O., and E.B.; Investigation, H.A., and R.A.; Methodology, H.A., and E.O.; Project administration, R.A.; Resources, E.O.; Software, H.A.; Visualization, M.O..; Writing-original draft preparation, E.O., and M.O.; Writing-review & editing, E.B. and E.O.; Supervision, M.O. All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTERESTS

The authors should confirm that there is no conflict of interest associated with this publication.

REFERENCES

1. International Energy Agency (IEA). Why AI and Energy Are the New Power Couple. IEA Commentary, 2 November 2023. Available online: https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple (accessed on 1 April 2025).
2. Capco. EU AI Act: Energy Implications. Capco Intelligence, 17 July 2024. Available online: https://www.capco.com/intelligence/capco-intelligence/eu-ai-act-energy-implications (accessed on 1 April March 2025).
3. PRISMA. PRISMA Extension for Scoping Reviews (PRISMA-ScR). PRISMA Statement, 2024. Available online: https://www.prisma-statement.org/scoping (accessed on 11 March 2025).
4. EUR-Lex Homepage. Available online: https://eur-lex.europa.eu/homepage.html (accessed on 1 April 2025).
5. Publications Office of the European Union. Available online: https://op.europa.eu/en/home (accessed on 1 April 2025).
6. European Union Agency for Cybersecurity (ENISA). Available online: https://www.enisa.europa.eu/ (accessed on 13 March 2025).

7. European Commission. Action Plan on the Digitalisation of the Energy Sector: Roadmap Launched. 2021. Available online: https://commission.europa.eu/news/action-plan-digitalisation-energy-sector-roadmap-launched-2021-07-27_en (accessed on 2 April 2025).
8. European Commission. New Network Code on Cybersecurity for the EU Electricity Sector. Energy-European Commission, 2024. Available online: https://energy.ec.europa.eu/news/new-network-code-cybersecurity-eu-electricity-sector-2024-03-11_en (accessed on 2 April 2025).
9. European Union Agency for Cybersecurity (ENISA). Cybersecurity and Privacy in AI: Forecasting Demand on Electricity Grids. ENISA Publications, 2023. Available online: https://www.enisa.europa.eu/publications/cybersecurity-and-privacy-in-ai-forecasting-demand-on-electricity-grids (accessed on 6 March 2025).
10. Katterbauer, K., Özbay, R.D., Yilmaz, S., Meral, G. 2030 On the Interface Between AI and Energy Regulations in China. *J. Recycl. Econ. Sustain. Policy* **2024**, 3, 51–62.
11. Morgan Lewis. The Intersection of Energy and Artificial Intelligence: Key Issues and Future Challenges. Morgan Lewis Publications, 12 August 2024. Available online: https://www.morganlewis.com/pubs/2024/08/the-intersection-of-energy-and-artificial-intelligence-key-issues-and-future-challenges (accessed on 1 April 2025).
12. European Commission. Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. COM/2021/206 Final, 2021. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689 (accessed on 4 February 2025).
13. European Parliament and Council. Regulation (EU) 2016/679—General Data Protection Regulation (GDPR). 2016. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 4 February 2025).
14. European Parliament and Council. Directive (EU) 2022/2555—NIS2 Directive. 2023. Available online: https://eur-lex.europa.eu/eli/dir/2022/2555 (accessed on 4 February 2025).
15. European Commission. Cyber Resilience Act (Proposal for a Regulation on Horizontal Cybersecurity Requirements for Products with Digital Elements). 2022. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0454 (accessed on 4 February 2025).
16. European Parliament and Council. Regulation (EU) 2019/881—Cybersecurity Act. 2019. Available online: https://eur-lex.europa.eu/eli/reg/2019/881/oj (accessed on 26 February 2025).
17. European Commission. Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. EUR-Lex, 2021. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206 (accessed on 2 April 2025).
18. European Union Agency for Cybersecurity (ENISA). Cybersecurity and Privacy in AI: Forecasting Demand on Electricity Grids. ENISA Publications, 2023. Available online: https://www.trustedai.ai/wp-content/uploads/2023/06/Cybersecurity-and-privacy-in-AI-Forecasting-demand-on-electricity-grids.pdf (accessed on 2 April 2025).
19. Integrate.ai. Improving Smart Grid Management with Federated Learning. Integrate.ai Blog, 12 February 2025. Available online: https://www.integrate.ai/blog/improving-smart-grid-management-with-federated-learning (accessed on 1 February 2025).

20. Center for Data Innovation. The Impact of the GDPR on Artificial Intelligence. Center for Data Innovation Report, 27 March 2018. Available online: https://www2.datainnovation.org/2018-impact-gdpr-ai.pdf (accessed on 1 February 2025).
21. European Commission. The NIS2 Directive: Strengthening Cybersecurity in the EU. Digital Strategy-European Commission, 18 October 2024. Available online: https://digital-strategy.ec.europa.eu/en/policies/nis2-directive (accessed on 2 February 2025).
22. KPMG. What Does NIS2 Mean for Energy Businesses? KPMG Insights, 19 April 2024. Available online: https://kpmg.com/uk/en/home/insights/2024/04/what-does-nis2-mean-for-energy-businesses.html (accessed on 2 March 2025).
23. Metomic. A Complete Guide to NIS2. Metomic Resource Centre, 21 November 2024. Available online: https://www.metomic.io/resource-centre/a-complete-guide-to-nis2 (accessed on 2 March 2025).
24. Norton Rose Fulbright. MiFID II: Frequency and Algorithmic Trading Obligations. Available online: https://www.nortonrosefulbright.com/en/knowledge/publications/6d7b8497/mifid-ii-mifir-series (accessed on 12 March 2025).
25. Financial Conduct Authority. MAR 5A.5 Systems and Controls for Algorithmic Trading. Available online: https://www.handbook.fca.org.uk/handbook/MAR/5A/5.html (accessed on 4 February 2025).
26. Advisera. Who Does NIS2 Apply to? Advisera Articles, 2024. Available online: https://advisera.com/articles/who-does-nis2-apply-to (accessed on 4 February 2025).
27. European Commission. Proposal for a Directive on Measures for a High Common Level of Cybersecurity Across the Union (NIS2 Directive). Available online: https://digital-strategy.ec.europa.eu/en/library/proposal-directive-measures-high-common-level-cybersecurity-across-union (accessed on 11 February 2025).
28. European Union. Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on Digital Operational Resilience for the Financial Sector (DORA). Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2554 (accessed on 4 April 2025).
29. Agency for the Cooperation of Energy Regulators (ACER). REMIT Quarterly Report—Monitoring Trading Patterns in Energy Markets. Available online: https://www.acer.europa.eu/sites/default/files/REMIT/REMIT%20Reports%20and%20Recommendations/REMIT%20Quarterly/REMITQuarterly_Q2_2020_3.0.pdf (accessed on 5 April 2025).
30. European Parliament and Council. Regulation (EU) No 1227/2011 on Wholesale Energy Market Integrity and Transparency (REMIT). Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011R1227 (accessed on 5 April 2025).
31. NERC (North American Electric Reliability Corporation). NERC CIP Standards—Critical Infrastructure Protection. 2023. Available online: https://www.nerc.com/pa/Stand/Pages/Project-2014-XX-Critical-Infrastructure-Protection-Version-5-Revisions.aspx (accessed on 7 April 2025).
32. U.S. Department of Energy (DOE). AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy; U.S. Department of Energy: Washington, DC, USA, 2024. Available online: https://www.energy.gov/sites/default/files/2024-

04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf (accessed on 2 February 2025).

33. National Institute of Standards and Technology (NIST). AI Risk Management Framework (AI RMF). 2023. Available online: https://www.nist.gov/itl/ai-risk-management-framework (accessed on 26 February 2025).
34. Office of Science and Technology Policy (OSTP). Blueprint for an AI Bill of Rights; The White House: Washington, DC, USA, 2022. Available online: https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ (accessed on 6 February 2025).
35. Executive Office of the President. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. Available online: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence (accessed on 16 February 2025).
36. Federal Trade Commission (FTC). On Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. 2024. Available online: https://www.ftc.gov/system/files/ftc_gov/pdf/FTC-AI-Use-Policy.pdf (accessed on 22 February 2025).
37. Electric Power Research Institute (EPRI). Advancing AI Integration in the European Energy Sector. EPRI Press Releases, 7 November 2024. Available online: https://europe.epri.com/press-releases/advancing-ai-integration-european-energy-sector (accessed on 4 January 2025).
38. Cyberspace Administration of China (CAC). Internet Information Service Algorithmic Recommendation Management Provisions; CAC: Beijing, China, 2022. Available online: https://www.chinalawtranslate.com/en/algorithms/ (accessed on 1 March 2025).
39. Cyberspace Administration of China (CAC). Provisions on the Administration of Deep Synthesis Internet Information Services; CAC: Beijing, China, 2023. Available online: https://www.chinalawtranslate.com/en/deep-synthesis/ (accessed on 3 March 2025).
40. Latham & Watkins LLP. China's New AI Regulations: Interim Measures for the Management of Generative AI Services; Latham & Watkins LLP: Beijing, China, 2023. Available online: https://www.lw.com/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf (accessed on 2 January 2025).
41. Ministry of Science and Technology of China (MOST). Ethical Norms for New AI Technologies; MOST: Beijing, China, 2021. Available online: https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/ (accessed on 1 January 2025).
42. National People's Congress of China. Personal Information Protection Law (PIPL). 2021. Available online: https://personalinformationprotectionlaw.com/ (accessed on 2 January 2025).
43. National People's Congress of China. Data Security Law (DSL). 2021. Available online: https://npcobserver.com/legislation/data-security-law/ (accessed on 10 January 2025).
44. Volkova, A.; Hatamian, M.; Anapyanova, A.; de Meer, H. Being Accountable is Smart: Navigating the Technical and Regulatory Landscape of AI-Based Services for Power Grid. arXiv 2024, arXiv:2408.01121v1. Available online: https://arxiv.org/abs/2408.01121 (accessed on 4 January 2025).
45. Artificial Intelligence Act. Article 112: Evaluation and Review. Artificial Intelligence Act Website, 2024. Available online: https://artificialintelligenceact.eu/article/112 (accessed on 1 January 2025).

46. European Consumer Organisation (BEUC). AI and Generative AI: Trilogue negotiations for the AI Act. 2023. Available online: https://www.beuc.eu/position-papers/ai-and-generative-ai-trilogue-negotiations-ai-act (accessed on 2 April 2025).

47. Elikplim, E. Energy Efficiency Dataset. Kaggle Datasets, 2019. Available online: https://www.kaggle.com/datasets/elikplim/eergy-efficiency-dataset

48. Ahmad, W. Unveiling Anomalies: Leveraging Machine Learning for Internal User Behaviour Analysis – Top 10 Use Cases. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, 8(1), 272–293.

49. Elias, A.H., Khairi, F.A., & Elias, A.H. Hybrid Machine-Learning Framework for Predicting Student Placement. *Journal of Transactions in Systems Engineering*, **2025,** 3(2), 403–419.

50. Durmuş, B., & Öznur İ.G. An application with meta-methods (MetaRF) based on random forest classifier. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2024**, 7(3), 80–97.

51. Shyam, K.M., Surapaneni, S., Dedeepya, P., Chowdary, N.S., & Thati, B. Enhancing Human Activity Recognition through Machine Learning Models: A Comparative Study. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, 8(1), 258–271.

52. Alkattan, H., Abdulkhaleq Noaman, S., Subhi Alhumaima, A., Al-Mahdawi, H., Abotaleb, M., & M. Mijwil, M. A Fusion-Based Machine Learning Framework for Lung Cancer Survival Prediction Using Clinical and Lifestyle Data. *Journal of Transactions in Systems Engineering*, **2025**, 3(2), 382–402.

53. Biswas, J., Mijwil, M.M., Farhan, L., Alkattan, H. Quantum computing in risk management and supply chain resilience. In Quantum Computing and Artificial Intelligence in Logistics and Supply Chain Management; Dutta, P.K., Bhattacharya, P., Verma, J.P., Chopra, A., Kundu, N.K., Aurangzeb, K., Eds.; Chapman and Hall/CRC: New York, NY, USA, **2025**, pp. 1-10.

54. Majumder, S., Alkattan, H., Dastjerdi, H.R. Quantum-blockchain scheme in fraud detection and risk management resilience. In Quantum Computing and Artificial Intelligence in Logistics and Supply Chain Management; Dutta, P.K., Bhattacharya, P., Verma, J.P., Chopra, A., Kundu, N.K., Aurangzeb, K., Eds.; Chapman and Hall/CRC: New York, NY, USA, **2025**, pp. 1-11.

55. Almusallam, N., Al Nuaimi, B.T., Alkattan, H., Abotaleb, M., Dhoska, K. Physics-Informed Neural Networks for Solving the Heat Equation in Thermal Engineering. *Int. J. Tech. Phys. Probl. Eng.* **2025**, 17(1), 375–382.

56. Alkattan, H., Subhi, A.A., Farhan, L., Al-mashhadani, G. Hybrid Model for Forecasting Temperature in Khartoum Based on CRU Data. *Mesopotamian J. Big Data* **2024**, 2024, 164–174

57. Alkattan, H., Abbas, N.R., Adelaja, O.A., Abotaleb, M., Ali, G. Data Mining Utilizing Various Leveled Clustering Procedures on the Position of Workers in a Data Innovation Firm. *Mesopotamian J. Comput. Sci.* **2024**, 2024, 104–109,

58. Alkattan, H., Abdullaev, S. Monitoring Wetlands in Southern Iraq Based on Landsat Data. Recent Advances in Environmental Science from the Euro-Mediterranean and Surrounding Regions (3rd Ed.); Advances in Science, Technology & Innovation (ASTI) book series; Springer: Cham, Switzerland, **2024**, pp. 433–436.