2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

COMPARATIVE MACHINE LEARNING AND DEEP LEARNING STUDY OF ENERGY PREDICTIONS IN URBAN AND RURAL BUILDINGS

Sharon Coffee[1], James O'Donnell[1, 2] and Cathal Hoare[1]
[1]School of Mechanical and Materials Engineering, University College Dublin, Ireland
UCD Energy Institute, University College Dublin, Ireland
[2]School of Engineering, University of Limerick, Limerick, Ireland

Abstract

The European Union's (EU) energy targets highlight the importance of retrofitting older buildings to reduce carbon emissions.-However, many rural properties remain in the lowest energy rating categories, complicating retrofitting efforts. Urban buildings dominate Energy Performance Certificates (EPC) models, while rural structures require tailored approaches due to their diversity and lower energy performance. This research compares machine and deep learning models to address gaps in predictive accuracy and scalability in retrofitting simulations. The methodology predicts EPC ratings based on renovation policies and improves regional segmentation and archetype classifications. These strategies offer insights for rural residential buildings aligned with EU energy efficiency standards.

Introduction

Meeting EU energy efficiency standards has become increasingly challenging for homeowners and policymakers. Retrofitting initiatives predominantly focused on urban areas where numerous properties have already achieved high-scoring EPC, frequently attaining A-standard ratings (Ali et al., 2019, 2020a, 2021, 2024a, 2024b; Djenouri et al., 2020; Hoare et al., 2022, 2023). These advancements have been supported by extensive retrofitting efforts and new builds designed to meet stringent efficiency standards. In contrast, rural areas face a significant disparity in retrofitting progress, with many properties still relying on fossil fuel heating systems and falling within the lowest energy rating bands, often requiring substantial interventions to meet EU targets (European Environment Agency, 2023). In Ireland, for example, approximately one-third of the residential building stock is in rural areas. Despite their disproportionately poor energy performance, these properties are underrepresented in energy modelling efforts.

Rural buildings, which frequently feature older construction and diverse structural characteristics, pose unique challenges to retrofitting efforts (Ali et al., 2024b). These areas are typically home to retired individuals or farming families with larger households, resulting in distinct energy consumption patterns compared to urban settings. Renovating rural homes in Europe presents added complexities as many rural properties rely on fossil fuel boilers such as gas and kerosene (Escribe and Vivier, 2025). These boilers contribute significantly to CO_2 emissions, further underscoring the need for targeted retrofitting strategies. Consequently, effective retrofitting strategies must prioritise improving the thermal efficiency of buildings through enhanced U-values (W/m^2K) for the building fabric and air tightness to achieve significant energy savings while reducing reliance on carbon-intensive heating systems, aligning with structured retrofit methodologies proposed by Ma et al. (2012). Such targeted renovation strategies are critical to addressing the urban-rural divide in achieving EU energy efficiency goals (Escribe and Vivier, 2025; European Environment Agency, 2023).

Traditional predictive models often oversimplify energy classifications, focusing on broader EPC rate bands (e.g. A, B, C, etc.) primarily in urban contexts (Ali et al., 2020, 2021, 2024a). This lack of granularity reduces the predictive power of these models, particularly for diverse rural building archetypes. Moreover, some conventional ML techniques often struggle to capture the complexities of energy performance modelling, limiting their scalability and effectiveness.

This research aims to address the gap in predictive accuracy and scalability of retrofitting simulations using advanced ML and DL methodologies. It evaluates and compares different ML and DL methodologies to assess their accuracy, scalability and performance in predicting EPC ratings for policy-driven renovation scenarios for rural and urban areas. This research examines how cutting-edge ML and DL models can boost EPC prediction accuracy while increasing the speed and scalability of retrofitting simulations across diverse building types and regions. The hypothesis suggests these methodologies will yield faster, more accurate predictions of EPC ratings, enabling scalable energy efficiency solutions at national and regional levels.

This study conducts a targeted analysis of retrofit impacts by categorising the EPC dataset into rural and urban sub-datasets. It introduces an enhanced predictive framework with 11 unique building archetypes, such as ground-floor, mid-floor and top-floor apartments; basements; mid-terrace and end-of-terrace houses; semi-detached and detached houses; bungalows; and maisonettes, surpassing previous models that used only four broad archetypes (Ali et al., 2020, 2021, 2024a; Hoare et al., 2023), thus improving energy performance assessments. New archetypes also include energy rating sub-bands (e.g., A1, A2, A3, etc.), enhancing the evaluation of retrofit scenarios.

Advanced DL techniques, like Neural Networks, Transformers and Autoencoders, are employed alongside Random Forest Regression to improve prediction accuracy and speed-up simulations. The methodology includes data preprocessing, model training, validation, and scenario-based retrofitting simulations tailored to diverse building contexts. A case study reviews ML/DL model performance in rural and urban settings, revealing how regional differences affect predictive accuracy, retrofit results and computational performance. This study offers actionable solutions for scalable, data-driven retrofitting strategies aligned with EU energy efficiency standards by addressing these challenges.

Related Research

This study builds on existing work in building energy performance modelling and ML applications for energy efficiency. A systematic review identified key studies focusing on urban contexts, archetype development and older modelling techniques. These works provide a foundation for understanding the strengths and limitations of current approaches.

Previous work

Several studies have advanced the field by applying ML techniques to energy retrofit predictions. Ali et al. (2020, 2024a) employed supervised and unsupervised ML models for urban residential buildings, optimising energy retrofit decisions through parametric simulations. However, these studies focus on broad urban archetypes and neglect rural contexts. Ali et al. (2021) categorised urban energy modelling approaches into top-down, bottom-up and hybrid frameworks, highlighting the scalability of urban building energy modelling (UBEM). Yet this scalability has not been extended to rural settings or more granular archetype classifications necessary for accurate retrofitting strategies. Hoare et al. (2023) proposed multi-scale archetypes limited to four dwelling types (terraced, detached, semi-detached and bungalows). This broad classification lacks detailed sub-band distinctions (e.g., A1, A2) for precise energy performance modelling. Similarly, Ali et al. (2019) introduced a data-driven framework incorporating geographic and demographic segmentation, but still relies on generalised archetypes. Other works, such as Hoare et al. (2022, 2023), leveraged linked data and digital twins to improve national-scale building stock modelling. While these methods enhance data integration and policymaker-centric tools, they offer limited direct applications for homeowners. Djenouri et al. (2020) also provided a taxonomy of ML techniques for smart buildings, emphasising theoretical frameworks over practical retrofitting applications. Collectively, these studies highlight key advancements but reveal critical gaps. Most research emphasises urban settings, neglecting rural properties, which differ significantly in construction type, age and energy consumption patterns. Another area for improvement lies in the broad classification of building archetypes. Existing models tend to oversimplify archetypes by grouping dwellings into general categories like detached houses, semi-detached houses, apartments, and terraced houses, which lack the granularity needed for precise energy performance modelling.

Furthermore, while ML techniques are widely applied, advanced DL models such as Transformers and Autoencoders remain underexplored in this domain. Many studies prioritise theoretical frameworks over the development of practical tools tailored to stakeholders, such as homeowners and policymakers.

Contributions

The study bridges theoretical advancements with practical applications by addressing this question, delivering a robust framework for scalable, data-driven energy efficiency solutions at national and regional levels.

This study addresses these gaps by introducing several novel contributions to the field of building energy performance modelling. First, the research investigates the advantages of using a more granular archetype classification by developing eleven distinct dwelling archetypes from a national EPC dataset. While national and regional factors may influence outcomes, this study specifically reports within the context of Irish EPC data. These archetypes include ground-floor, mid-floor and top-floor apartments; basements; mid-terrace and end-of-terrace houses; semi-detached houses; detached houses; bungalows; and maisonettes. Additionally, sub-rate band classifications, such as A1, A2, A3, etc., are incorporated to improve the precision of energy performance evaluations. This level of detail surpasses previous studies limited to four broad categories, enabling more accurate predictions of retrofit outcomes.

Second, the study incorporates regional segmentation to differentiate between rural and urban contexts. While the EPC rating system remains consistent across all residential properties, this research introduces a novel subdivision of the dataset into rural and urban categories. This distinction allows for a more nuanced analysis of how building location and context influence energy performance, retrofit potential and model accuracy. This targeted approach addresses rural buildings' specific retrofit challenges, ensuring the framework applies to diverse geographic settings. This dual focus bridges a critical gap in current research and enhances the versatility of the modelling framework.

Third, the study employs advanced ML and DL techniques to improve prediction accuracy and scalability. Random Forest Regression is used as a baseline model for efficiency and interpretability (Ali et al., 2020). Additionally, the research explores Neural Networks, Autoencoders, and Transformers to capture complex patterns in the data and evaluate their effectiveness in predicting EPC ratings. These techniques are assessed for their ability to capture complex relationships in the data, improve computational performance and generalise across different building archetypes and regions.

Finally, this study offers practical insights into retrofitting by integrating detailed archetypes with advanced models. The outcomes provide actionable insights that could inform the development of practical tools, such as retrofit decision-support systems or policy guidelines, aligned with national and EU energy efficiency standards.

Methodology

The study follows a systematic workflow (Figure 1) for processing energy performance data, training ML and DL models, simulating retrofitting scenarios and evaluating model performance. The workflow consists of four primary stages: 1) Data Preprocessing, 2) ML/DL Model Training, 3) Retrofitting Simulation and Prediction, and 4) Model Comparison.

This methodology supports a structured comparison of advanced ML and DL techniques for predicting energy performance outcomes under retrofit scenarios. By using a nationally representative EPC dataset, the workflow enables a controlled environment for evaluating how various models perform across different dwelling types and regional settings. The approach integrates data-driven feature selection, building archetype classification and regional segmentation to better capture the diversity of the residential building stock. By combining these elements, the methodology offers a scalable and replicable framework that supports the development of targeted retrofit strategies grounded in empirical data.

Each methodology stage is purposefully aligned to ensure traceability, consistency and model comparability. The simulation framework applies uniform criteria across all models to assess accuracy, scalability and generalisation. This enables a robust cross-model and urban-rural comparison, revealing how regional and structural factors influence model performance.

Data Preprocessing

The anonymised Energy Performance Certificate (EPC) dataset is the primary data source of detailed energy performance information. From this dataset, 11 distinct building archetypes were developed to capture the diversity of housing structures, including ground-floor, mid-floor, and top-floor apartments; basements; mid-terrace and end-of-terrace houses; semi-detached and detached houses; bungalows; and maisonettes. These archetypes align with the dwelling type categories recorded in the EPC dataset. While previous studies consolidated these into four broader archetypes (apartments, terraced houses, semi-detached houses and detached houses), this study retains all 11 to preserve granularity and enables more precise modelling of retrofit impacts across varied dwelling types. Average U-values for windows, doors, floors, and roofs were calculated for each archetype and energy sub-rate band (e.g., A1, A2, etc.) from the EPC dataset, providing a basis for parametric simulations and predictive modelling, consistent with BIM-based methodologies proposed in retrofit literature (Sanhudo et al., 2018). To ensure data quality, model interpretability and computational efficiency, the original datasets containing 212 columns were reduced to 106 columns. Features with excessive missing values were removed, following standard practices in data-driven energy modelling to minimise bias and prevent unreliable imputation. Highly correlated variables that provided redundant information were eliminated to maintain independent and meaningful predictors. To further enhance dataset reliability, data inconsistencies, such as trailing spaces, inconsistent formatting and erroneous entries, were corrected. Each building was assigned a unique property ID to facilitate traceability and maintain data integrity across different modelling stages. Additionally, geographical segmentation was applied to categorise properties by province and urban/rural contexts, ensuring that regional energy performance variations could be effectively analysed. These preprocessing steps refined the dataset to include only relevant, high-quality features, improving model accuracy while maintaining sufficient data coverage for robust predictions.

ML/DL Model Training

In the second stage, the cleaned datasets (building stock synthetic data) were split into training, validation and test

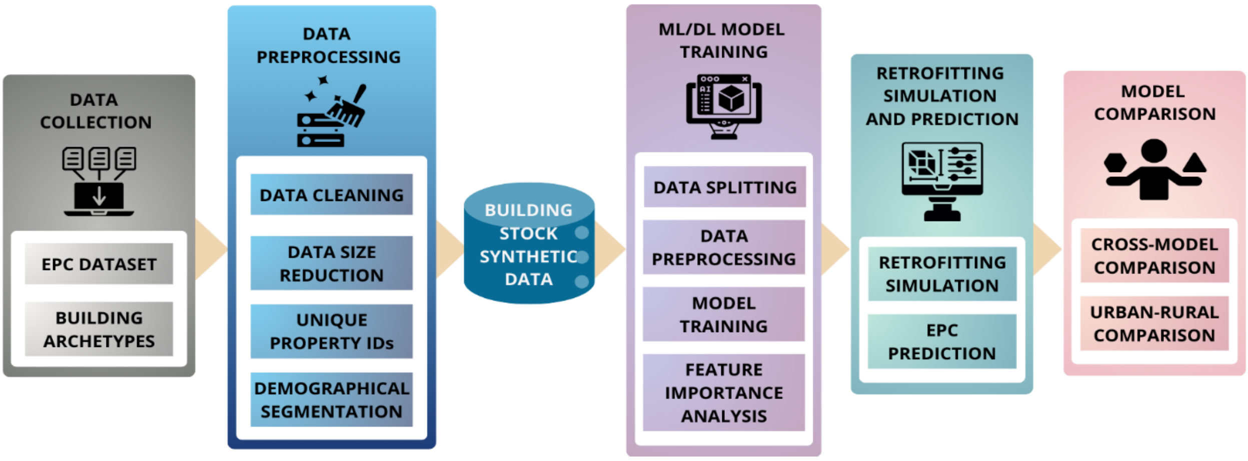


Figure 1: Methodology for Data-Driven EPC Prediction and Retrofitting Simulation Using ML and DL Techniques

sets using the hold-out method in a 60/20/20 split. The hold-out method was selected because it is well-suited to larger datasets (Yadav and Shukla, 2016), ensuring unbiased model evaluation by keeping the test set separate during development. The 60/20/20 split was chosen as it provides sufficient training data while maintaining a significant portion for validation and testing, ensuring robust model assessment across different data subsets. The same shuffled datasets were used across all models to ensure fair performance comparisons.

Before training, preprocessing steps included imputing missing values to maintain dataset completeness, scaling numeric features for uniformity and encoding categorical features for compatibility with ML and DL algorithms. The models evaluated include one traditional ML model (Random Forest Regression) and three DL models: Neural Networks, Transformers and Autoencoders. Random Forest Regression was chosen as the baseline model due to its efficiency and interpretability (Ali et al., 2020); Neural Networks were applied for their capacity to capture nonlinear relationships; Transformers for their scalability in processing high-dimensional datasets; and Autoencoders for feature extraction and dimensionality reduction. Only four models were trained, but the Autoencoder was assessed in two configurations (Simplified and Complex with Batch Normalisation) to evaluate architecture complexity. Consequently, five models were presented in the Case Study. Feature selection aimed to identify key parameters for better computational efficiency and predictive accuracy. Priority was given to thermally significant parameters affecting EPC ratings, including U-values, heating system efficiencies, and insulation types. While the selected features differ slightly from those used in previous studies, which reduced inputs to a few core (Ali et al., 2024a; Egan et al., 2018), this study applied model-driven selection techniques. These included the built-in feature importance mechanism from Random Forest Regression and SHAP (Shapley Additive exPlanations) analysis for DL models. Notably, many top-ranked features aligned with those highlighted in prior work, reinforcing the importance of thermal and structural attributes in residential energy performance prediction.

Retrofitting Simulation and Prediction

The retrofitting simulation used average U-values derived from the 11 archetypes designed to preserve dwelling-type granularity in the EPC dataset. This enabled the simulation of more realistic and precise energy efficiency improvements across diverse building types and retrofit scenarios. ML and DL models predicted new EPC ratings for retrofitted properties by adjusting key building parameters, such as walls, roofs, windows, floors, and doors. The study focused on lower EPC rating bands (E1, E2, F, and G) in retrofitting simulations, as these buildings require urgent upgrades. This aligns with the EU's goal to phase out fossil fuel use in buildings, requiring all new buildings to meet zero-emission standards by 2030 and progressively transforming existing buildings to zero-emission performance by 2050 (European Parliament and Council of the European Union., 2024). With nearly one in five rural properties in these lowest EPC bands, study findings can inform targeted retrofit strategies that ensure energy savings, cost-effectiveness, and feasibility. These simulations reveal energy savings and carbon reduction opportunities across archetypes and regions, aiding in identifying effective retrofitting strategies.

Model Comparison

The final stage compared model performance from two perspectives: cross-model and urban-rural comparisons. Models were evaluated on accuracy, computational efficiency, and generalisation across diverse archetypes and scenarios. Model performance was assessed using Root Mean Squared Error (RMSE) and R^2 metrics to evaluate predictive accuracy and generalisation capabilities. The Case Study section provides a detailed comparison of these results, including cross-model and urban-rural analysis.

Rural regions, characterised by older, structurally diverse buildings, presented distinct challenges compared to urban areas, where building stock reflects more homogeneous construction and retrofitting efforts. This segmentation provided valuable insights into how regional differences affect predictive accuracy and computational performance, guiding stakeholders in selecting suitable models based on specific needs.

Case Study

This case study evaluates the performance of five ML and DL models for predicting residential buildings' energy performance: Random Forest Regression, Neural Network, Transformer and two Autoencoder variants. Four distinct model types were trained, with the Autoencoder evaluated in Simplified and Complex configurations to examine the impact of architectural depth. The Complex Autoencoder included Batch Normalisation, which is absent in the Simplified model. The analysis uses two sub-datasets derived from the anonymised dataset made available by the Sustainable Energy Authority of Ireland (SEAI)[1]. These sub-datasets of Irish residential buildings represent counties (rural areas) and cities (urban areas), excluding Dublin City and Dublin County, providing a comparative evaluation of model performance across distinct regional contexts.

Overview of Sub-datasets

Figure 2 represents a comparative overview of the energy rating distribution across four sub-datasets: Counties, Cities, Dublin County and Dublin City. The Counties sub-dataset, which predominantly represents rural areas, includes 784,716 properties, with a significant number showing poor energy performance. Specifically, 148,609 (18.94%) fall within the lowest rating bands (E1 to G), while 236,283 (30.11%) are in the higher performance categories (A1 to B3). In contrast, the Cities sub-dataset includes 76,399 properties, of which 14,264 (18.67%) fall

---

[1] https://ndber.seai.ie/berresearchtool/ber/search.aspx

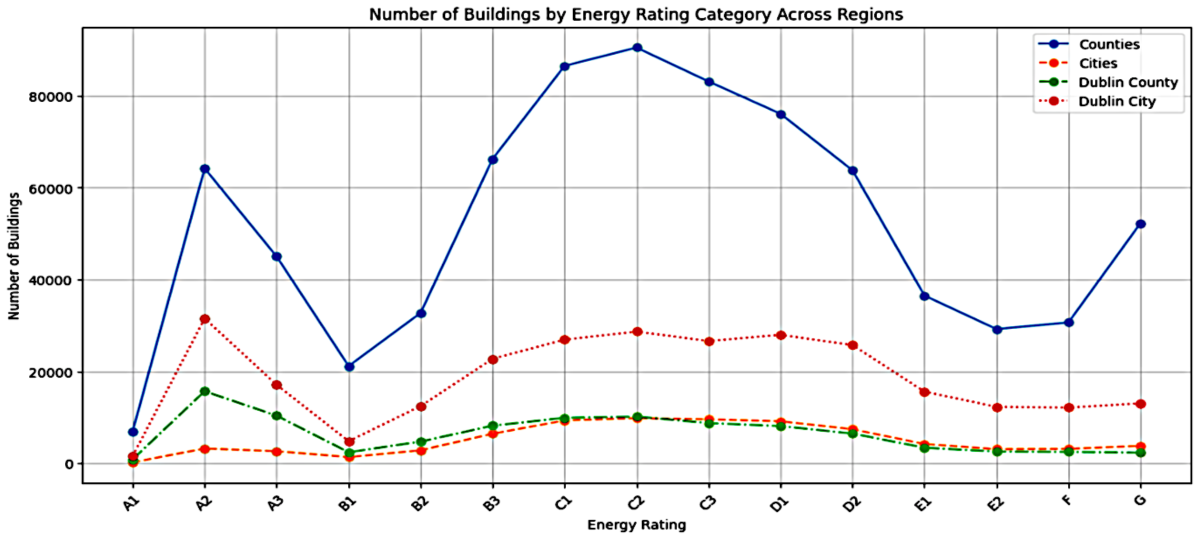


Figure 2: Comparison of the number of buildings by energy rating category across four sub-datasets (Counties, Cities, Dublin County and Dublin City)

into the lower E1 to G range, and 16,788 (21.97%) are rated A1 to B3. The Dublin County sub-dataset contains 96,612 properties, with only 10,839 (11.22%) falling into the lowest rating bands and 42,296 (43.78%) achieving the highest energy performance levels. Dublin City, with 279,267 properties, shows similar trends, with 53,119 (19.02%) in the E1 to G bands and 90,226 (32.31%) in the A1 to B3 range.

This visual comparison highlights the urban-rural divide in energy performance. Urban areas such as Dublin tend to have a higher proportion of energy-efficient buildings due to more widespread retrofitting and newer construction. In contrast, rural areas continue to face significant challenges, including a predominance of older buildings and reliance on fossil-fuel heating systems such as oil-fired boilers, which contribute disproportionately to CO_2 emissions. This disparity reinforces the importance of targeted retrofitting strategies that prioritise improving thermal efficiency (through lower U-values) and replacing carbon-intensive heating systems (Escribe and Vivier, 2025; European Environment Agency, 2023). This pattern aligns with previous studies that reported significant retrofitting progress in urban areas, but highlighted the underrepresentation of rural contexts in energy modelling (Hoare et al., 2023). The sub-datasets offer a valuable foundation for evaluating the performance of ML and DL models in bridging these energy performance gaps and informing policy-driven renovation strategies across diverse regional contexts.

Feature Selection and Importance Analysis

Feature importance analysis was performed using built-in importance for Random Forest Regression and SHAP for Neural Network, Transformer and Autoencoder models to validate and interpret the trained models. Although feature selection occurred during preprocessing, this step assessed how each model prioritised inputs during training. The analysis showed that thermal properties (U-values), heating system efficiencies and structural details (dwelling type, total floor area) consistently ranked as the most influential variables across both ML and DL models. These findings support previous urban energy modelling studies (Ali et al., 2024a), highlighting the importance of thermal and structural attributes for accurate energy predictions. Table 1 shows the feature set used in all models, confirming the alignment of data insights with expectations.

Table 1: List of Streamlined Features used in model training

Streamlined Features		
Structural Details	**Thermal Properties**	**Heating System Efficiencies**
Property ID	U-value Wall	HS Main System Efficiencies
County Name	U-value Roof	HS Supply System
Dwelling Type	U-value Floor	WH Main System Efficiencies
Construction Year	U-value Window	Solar Hot Water Heating
Rating Type	U-value Door	Insulation Type
No. Storeys	Ground Floor U-value	Insulation Thickness
Total Floor Area		
Assessment Date		

Model Performance Metrics

The models were evaluated using Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) to assess accuracy and generalisation capabilities. RMSE measures the average magnitude of prediction errors, while R^2 evaluates the proportion of variance explained by the model. Previous studies predominantly applied these metrics in urban contexts, revealing their utility for large-scale predictive modelling (Ali et al., 2020, 2021, 2024a). For brevity, detailed formulas and explanations of these metrics are cited in previous works, as they have been extensively discussed (Ali et al., 2024a). The focus here is to highlight how these metrics demonstrate the comparative advantages of advanced DL models over traditional ML approaches.

Various Operating Systems for Computational Performance

The models were tested across three operating systems – Google Colab with 12GB RAM, Windows Server OS with 128GB RAM, and MacOS with 64GB RAM - to evaluate the impact of computational resources on training times and scalability. This approach builds on urban energy modelling methodologies that emphasise scalability as a critical factor in deploying predictive frameworks (Ali et al., 2024a). While resource-intensive models like Transformers and Autoencoders encountered computational challenges in environments with limited memory, Random Forest Regression demonstrated robust performance across all platforms, underscoring its efficiency for practical applications. The findings provide actionable insights into selecting computational environments based on data complexity and model requirements, offering scalability for large-scale, policy-driven retrofitting scenarios.

Contextual Comparisons with Previous Research

This study extends the work of Ali et al. (2024a, 2024b) by incorporating rural sub-datasets and evaluating advanced DL models. While prior works primarily focused on urban areas and highlighted the scalability of ML models, this research expands the scope by addressing the distinct challenges of rural regions, such as older building stock and carbon-intensive heating systems. Using eleven detailed archetypes, as opposed to four broad categories in earlier studies, enhances predictive granularity and aligns with contemporary best practices in energy modelling (Ali et al., 2019, 2024a; Hoare et al., 2023).

Results

The models were evaluated on the Counties (rural) and Cities (urban) sub-datasets to assess their robustness, consistency and ability to generalise across diverse building archetypes and geographic contexts. The analysis provided insights into how advanced DL models performed relative to simpler ML models in predicting EPC ratings, particularly under different data characteristics.

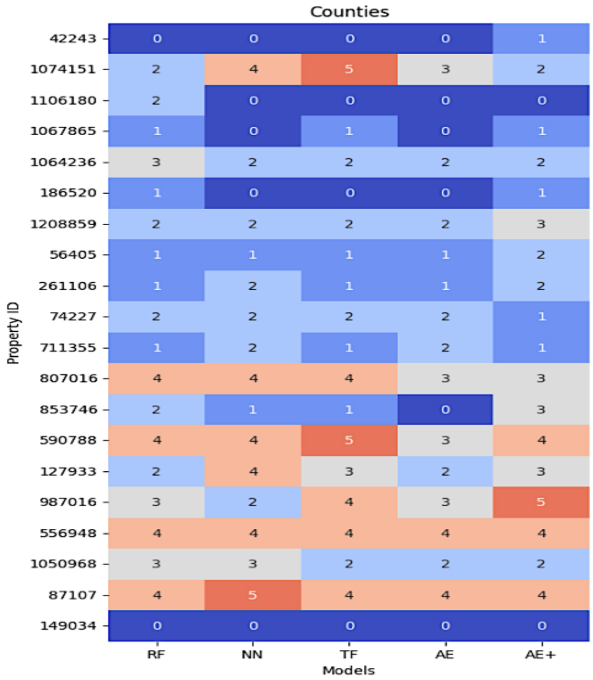
Performance Across Sub-datasets

In the Counties sub-dataset, representing rural areas with 784,716 properties, the models faced challenges due to the structural diversity and variability in retrofitting levels. Advanced DL models, such as Transformers and Autoencoders, struggled to maintain accuracy compared to Random Forest Regression and Neural Networks. This trend was particularly evident in heatmap analyses (Figures 3 and 4), where the DL models showed a higher proportion of deviations from actual EPC ratings. For instance, Autoencoders demonstrated higher prediction variability and performed slightly worse in capturing the energy rating trends for rural properties, reflecting the impact of noise and structural heterogeneity.

Conversely, the Cities sub-dataset, comprising 76,399 properties, revealed improved performance for DL models. Transformers and Autoencoders achieved better predictive accuracy and reduced deviations from actual EPC ratings. This suggests that the relatively uniform construction practices and retrofitting measures in urban settings allowed these models to leverage their capacity to detect complex patterns more effectively. This improvement narrowed the performance gap between DL and simpler models like Random Forest Regression and Neural Networks.

Heatmap Analysis

Heatmap analyses revealed key differences in model performance across sub-datasets.

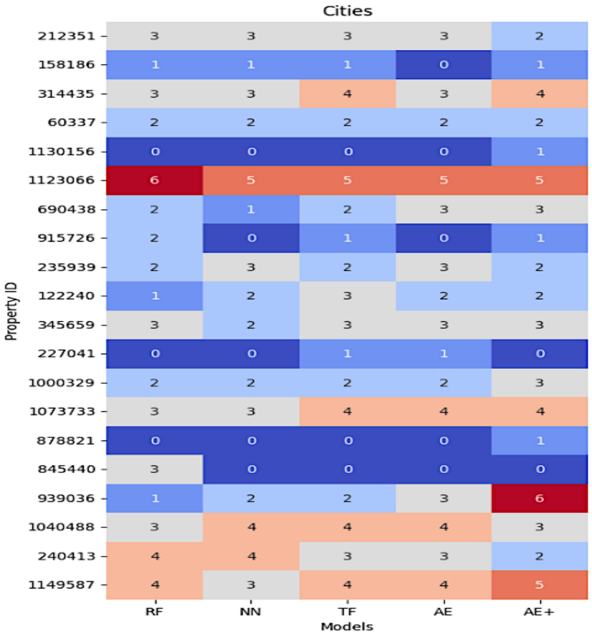


Property ID	RF	NN	TF	AE	AE+
42243	0	0	0	0	1
1074151	2	4	5	3	2
1106180	2	0	0	0	0
1067865	1	0	1	0	1
1064236	3	2	2	2	2
186520	1	0	0	0	1
1208859	2	2	2	2	3
56405	1	1	1	1	2
261106	1	2	1	1	2
74227	2	2	2	2	1
711355	1	2	1	2	1
807016	4	4	4	3	3
853746	2	1	1	0	3
590788	4	4	5	3	4
127933	2	4	3	2	3
987016	3	2	4	3	5
556948	4	4	4	4	4
1050968	3	3	2	2	2
87107	4	5	4	4	4
149034	0	0	0	0	0

Figure 3: Prediction deviations for the Counties sub-dataset

Figure 3 presents a heatmap of prediction deviations for the Counties sub-datasets across all five models. Random Forest Regression and Neural Networks demonstrated the most consistent performance, with 65% of their predictions falling within a 0-2 rating deviation from the actual EPC values. In contrast, DL models, particularly Transformers and both variants of the Autoencoder, exhibited more significant variability. For instance, Transformers achieved 45% accuracy within the stricter 0-1 deviation range, compared to 50% for Random Forest Regression. The complex Autoencoder showed the widest prediction spread, reflecting the challenges of generalising across older, structurally diverse rural building stock.

Figure 4 shows the corresponding heatmap for the Cities sub-dataset. Here, predictive accuracy improved for DL models. Transformers and Autoencoders performed comparably to Random Forest Regression and Neural Networks, with Transformers achieving 60% accuracy within the 0-2 deviation range. This performance boost is likely due to more uniform construction practices and consistent retrofitting efforts in urban areas, which enable DL models to learn and generalise patterns more effectively.



Property ID	RF	NN	TF	AE	AE+
212351	3	3	3	3	2
158186	1	1	1	0	1
314435	3	3	4	3	4
60337	2	2	2	2	2
1130156	0	0	0	0	1
1123066	6	5	5	5	5
690438	2	1	2	3	3
915726	2	0	1	0	1
235939	2	3	2	3	2
122240	1	2	3	2	2
345659	3	2	3	3	3
227041	0	0	1	1	0
1000329	2	2	2	2	3
1073733	3	3	4	4	4
878821	0	0	0	0	1
845440	3	0	0	0	0
939036	1	2	2	3	6
1040488	3	4	4	4	3
240413	4	4	3	3	2
1149587	4	3	4	4	5

Figure 4: Prediction deviations for the Cities sub-dataset

RMSE and R^2 Comparisons

The RMSE and R^2 metrics further confirmed the performance trends observed in the heatmap analyses. In the Counties sub-dataset, Random Forest Regression achieved the lowest test RMSE of 54.97 and the highest test R^2 of 0.891, indicating its strong predictive accuracy and generalisation capability in rural contexts. DL models like Transformers (Test RMSE: 58.87, Test R^2: 0.875) and Neural Networks (Test RMSE: 57.98, Test R^2: 0.879) exhibited slightly weaker performance, particularly in handling the variability and noise of rural data. The Autoencoder models showed poor training RMSE values (282.6 for the complex variant), likely due to their reconstruction-focused objective. However, they demonstrated significant improvements in validation and test RMSEs, highlighting their ability to generalise well despite challenges during training.

The performance gap between DL and ML models narrowed in the Cities sub-dataset, with Random Forest Regression still achieving the best results (Test RMSE: 39.54, Test R^2: 0.902). Transformers and Neural Networks delivered competitive performance, with test RMSE values of 46.65 and 45.40, respectively and R^2 scores above 0.86. The Autoencoder models showed consistent improvements in validation and test RMSE values, reflecting better generalisation in the more consistent urban sub-dataset.

Impact of Sub-dataset Size and Characteristics

The results underscore that sub-dataset size and the structural diversity and retrofitting disparities between rural and urban regions significantly affect the performance of advanced DL models. The Counties sub-dataset, while larger, contained greater heterogeneity in building types and retrofitting levels, posing challenges for DL models to generalise effectively. In contrast, the smaller Cities sub-dataset exhibited more homogeneous construction practices and energy performance trends, enabling DL models to leverage their capacity for capturing complex patterns. The Autoencoder models highlighted this trend. Despite high training RMSE values, their validation and test RMSEs improved substantially, indicating effective generalisation when applied to the urban sub-dataset. This discrepancy can be attributed to their reconstruction-focused training objective and the challenges of handling diverse rural data with higher noise levels.

Summary of Observations

Random Forest Regression consistently outperformed other models in both sub-datasets, demonstrating a reliable balance of accuracy, efficiency and generalisation. Neural Networks and Transformers excelled in urban contexts, but exhibited slightly weaker performance in rural areas due to greater variability in building archetypes and retrofitting levels. Despite challenges in training, Autoencoders displayed strong generalisation capabilities in validation and test phases, particularly in the Cities sub-dataset. Differences in sub-dataset characteristics, including size, structural diversity and retrofitting efforts, significantly influenced the effectiveness of ML and DL models, with advanced DL models benefiting more from sub-datasets with less variability.

Conclusions and Future Work

The study presents a novel framework that combines detailed building archetype classifications, regional segmentation, and advanced ML/DL models to enhance EPC prediction accuracy and scalability. The detailed archetype granularity, featuring 11 distinct archetypes and sub-rate bands, enables a more nuanced assessment of retrofitting impacts. By incorporating regional segmentation, the framework effectively addresses the unique challenges of rural and urban settings, providing actionable insights for energy performance modelling and policy-driven initiative strategies. The results underscore the utility of Random Forest Regression as a practical baseline model for large-scale EPC predictions while also exploring the potential of DL models like Transformers and Autoencoders to capture complex patterns, particularly in more consistent urban sub-datasets. Additionally, the study emphasises the importance of data preprocessing, including feature selection and geographical segmentation, in enhancing model performance.

This research enhances energy modelling with a scalable methodology that links data insights to retrofit planning.

By refining archetype classification and utilising advanced ML/DL models on EPC data in rural and urban areas, it offers a replicable framework for national and EU retrofitting strategies. Notably, regional segmentation enhances model accuracy, particularly for rural buildings. A key finding is that simpler models, like Random Forest Regression, can outperform complex DL architectures when data variability is high or when interpretability and efficiency matter, providing practical guidance for stakeholders managing performance and resources. DL models show promise in structured urban sub-datasets, but optimisation should be context-specific. In many cases, the marginal gains from tuning these models may not justify the added complexity and cost. Hybrid approaches, combining interpretable ML with DL techniques, may provide a more balanced solution.

Future work will expand the analysis to include CO_2 emissions predictions for a holistic retrofitting impact evaluation. This data integration will enhance understanding of energy performance improvements and environmental sustainability. Developing practical tools and user-friendly applications for policymakers and homeowners is a priority, enabling real-time energy assessments and decision-making. Efforts will also focus on optimising deep learning models for better performance and scalability, especially with datasets that exhibit high variability and noise, like rural building stocks. Exploring hybrid modelling that combines simpler ML models with advanced DL techniques may yield robust and scalable solutions. This research lays a foundation for advancements in data-driven energy performance modelling to support EU energy efficiency targets and provide effective retrofitting solutions for diverse building contexts.

Acknowledgements

This publication has emanated from research supported by Research Ireland through the US-Ireland R&D Partnership Research Grant 20/US/3695 and the European Union's Horizon 2020 Europe Research and Innovation Program (Grant Agreement No 101037075). The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Research Ireland, the European Union or other funding agencies.

References

Ali, U., Bano, S., Shamsi, M.H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., O'Donnell, J., 2024a. Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. Energy Build. 303, 113768.

Ali, U., Bano, S., Shamsi, M.H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., O'Donnell, J., 2024b. Urban residential building stock synthetic datasets for building energy performance analysis. Data Brief 53, 110241.

Ali, U., Shamsi, M.H., Bohacek, M., Hoare, C., Purcell, K., Mangina, E., O'Donnell, J., 2020. A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings. Appl. Energy 267, 114861.

Ali, U., Shamsi, M.H., Hoare, C., Mangina, E., O'Donnell, J., 2019. A data-driven approach for multi-scale building archetypes development. Energy Build. 202, 109364.

Ali, U., Shamsi, M.H., Hoare, C., Mangina, E., O'Donnell, J., 2021. Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. Energy Build. 246, 111073.

Djenouri, D., Laidi, R., Djenouri, Y., Balasingham, I., 2020. Machine Learning for Smart Building Applications: Review and Taxonomy. ACM Comput. Surv. 52, 1–36.

Egan, J., Finn, D., Deogene Soares, P.H., Rocha Baumann, V.A., Aghamolaei, R., Beagon, P., Neu, O., Pallonetto, F., O'Donnell, J., 2018. Definition of a useful minimal-set of accurately-specified input data for Building Energy Performance Simulation. Energy Build. 165, 172–183.

Escribe, C., Vivier, L., 2025. Banning new gas boilers as a no-regret mitigation option. Nat. Commun. 16, 49.

European Environment Agency, 2023. Decarbonising heating and cooling - a climate imperative.

European Parliament and Council of the European Union., 2024. Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings (recast). Off. J. Eur. Union L1275, 68.

Hoare, C., Aghamolaei, R., Lynch, M., Gaur, A., O'Donnell, J., 2022. A linked data approach to multi-scale energy modelling. Adv. Eng. Inform. 54, 101719.

Hoare, C., AlQazzaz, T.B.M., Ali, U., Hu, S., O'Donnell, J., 2023. Development of a National Scale Digital Twin for Domestic Building Stock. LDAC 12.

Ma, Z., Cooper, P., Daly, D., Ledo, L., 2012. Existing building retrofits: Methodology and state-of-the-art. Energy Build. 55, 889–902.

Sanhudo, L., Ramos, N.M.M., Poças Martins, J., Almeida, R.M.S.F., Barreira, E., Simões, M.L., Cardoso, V., 2018. Building information modeling for energy retrofitting – A review. Renew. Sustain. Energy Rev. 89, 249–260.

Yadav, S., Shukla, S., 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. Presented at the Proceedings - 6th International Advanced Computing Conference, IACC 2016, Institute of Electrical and Electronics Engineers Inc., pp. 78–83.