A Comprehensive Approach for Automated Anomaly Detection and Enhancement of EPC Datasets for Decarbonisation

Sahar Mirzaie and Zhihao Zhang

Institute for Sustainable Built Environment, Heriot-Watt University, UK
s.mirzaie@hw.ac.uk, zz5001@hw.ac.uk

Abstract

Energy Performance Certificates (EPCs) are pivotal for evaluating building energy efficiency, informing decarbonisation policies, and supporting retrofit. However, errors in EPC datasets compromise their reliability. This study develops an automated anomaly detection framework integrating Machine Learning (ML) and auxiliary datasets to enhance EPC data accuracy at scale. By reviewing validation methods and identifying gaps, this paper proposes a systematic framework for improving data quality assurance, reinforcing EPCs as a key resource for policy and decarbonisation efforts.

Keywords –

Energy Performance Certificates (EPCs), Machine Learning, Data Validation, Automating Quality Assurance, Decarbonisation

1 Introduction

Energy Performance Certificates (EPCs) have emerged as a cornerstone of energy efficiency policy across Europe [1], providing standardised metrics to evaluate the energy performance of buildings. Introduced in 2007, EPCs facilitate the monitoring of policy progress, guide retrofitting initiatives and inform decisions regarding financial incentives. In the United Kingdom (UK), an EPC is mandatory when a building is constructed, sold, or rented, are required for various funding schemes (e.g. the Green Deal) and remains valid for ten years. With 27 million domestic EPCs in the UK, these datasets offer key insights into national building energy performance [2].

In the UK, the Minimum Energy Efficiency Standards (MEES), introduced in 2018, prohibit private landlords from renting properties with an EPC rating of F or G, with the effectiveness of MEES heavily reliant on EPC accuracy [2,3]. Stock models use adapted EPC calculations to estimate input uncertainties and energy efficiency policy impact at scale [4].

Despite their extensive use, EPCs have faced criticism due to frequent errors and inconsistencies, such as discrepancies in assessments of the same property [3,5,6]. These issues undermine their credibility as policy tools and limit their application in areas such as retrofitting and carbon reduction planning. Addressing these challenges requires the implementation of rigorous validation and editing methodologies to ensure data accuracy and reliability.

Several studies have examined the accuracy of average assessment using the UK EPC dataset, as summarised in Table 1. They acknowledge that the results are not necessarily applicable to a single EPC. Yuan and Choudhary [3] compared the original and renewed EPCs of office buildings with small cooling and heating systems published between 2010 and 2020, looking at changed asset rating using clustering analysis. The non-domestic EPCs offer little information about the building’s physical conditions, therefore they assumed a change in the recommendations reflects a change in the building conditions. They conclude that buildings with unchanged conditions can obtain a decreased or increased EPC rating in subsequent assessments.

With residential buildings responsible for a significant share of carbon emissions and the majority of the 2050 building stock already in place, ensuring the accuracy and reliability of EPCs as a tool for evaluating progress for decarbonisation of the domestic building stock is vital. Therefore, this work focuses on domestic EPCs.

To produce an EPC for a residential building, certified energy assessors collect data and, in many cases, make assumptions about the building's energy-related features, such as construction materials, heating systems, and insulation. This information is processed using the Government's Standard Assessment Procedure (SAP) for new builds or the reduced Standard Assessment Procedure (rdSAP) for existing homes. The SAP algorithm models the building in a steady-state condition to calculate theoretical heat loss, using estimated thermal properties, standardised heating patterns and occupancy, and fixed fuel price assumptions. It then assigns an Energy Efficiency Rating (EER) on a 100-point scale, based on energy costs per square metre, along with an associated EPC band ranging from A (most efficient) to G (least efficient) [2].

SAP and rdSAP are set to be replaced by the Home Energy Model (HEM) in the second half of 2026. Unlike SAP's steady-state modelling approach, HEM will utilise dynamic modelling and updated methodologies to provide a more accurate representation of real-world energy usage, carbon emissions and heating demand [7]. Calculation methodologies change over time, and assessors use different software and have different approaches to assessment. All these cause variations in the results [5,8]. This study however is not concerned with the causes of these variations but rather identifying potential EPC errors for a dwelling. Comparison of the gap between the modelling results and actual energy use of the building is beyond the scope of this work. That said, similar to other studies, it is assumed that the dwelling's rating should not change in repeated assessments unless there has been a change to the fabric or systems, reflected in their recommendations [3,8].

This study evaluates existing EPC validation methods, focusing on their strengths and limitations, particularly for enabling automated EPC data applications. It explores the integration of auxiliary datasets and machine learning (ML) techniques to enhance the automation of validation processes and data quality. While much prior research has concentrated on repeated EPC assessments [3,5,8,9], this paper adopts a broader approach by examining variables and inputs involved in generating a single EPC as well. Through this wide-ranging analysis, this study proposes a comprehensive framework to improve automation of EPC data quality assurance, contributing to the enhancement of EPC reliability, thereby supporting their effective application in energy efficiency policies and broader research contexts.

2 Literature Review

To understand how EPCs have been validated a comprehensive literature review of Scopus database was conducted, using 'Energy Performance Certificate' or 'EPC' in the title, with a particular focus on domestic properties and validation techniques. Additional studies were incorporated using the 'snowball' approach. The search results indicated that relevant publications on this topic emerged from 2016 onwards.

Table 1. Overview of EPC Validation Studies in the UK: Key Characteristics, Approaches, and Analytical Methods

Ref.	Year	Country	Domestic	Sample Size	Sample Period	EPC values	Validation Approach	Validation levels	Statistical or Machine Learning Approach
[3]	2023	England and Wales	N	5400 simple office buildings	2010-2020	Asset Rating, Standard Emission Rate (SER), Building Emission Rate (BER), Recommendations	Changed asset rating using clustering analysis.	2	Local weighted polynomial regression to study the distribution. Clustering analysis for association pattern recognition: GMM.
[5]	2017	England and Wales	D	145 audits of 29 dwellings	2016	EER, age, construction, and property type, physical properties such as floor area, fabric and systems efficiency	At least 4 independent energy assessments	3	-
[8]	2019	England and Wales	D	1.6 million EPCs from 13 million dwellings	2009-2016	EER	Repeated EPC assessments and change in EER. Exploratory data analysis to decide how to apply this model to available data.	2	A semi-parametric statistical model of how measurement error contributes to variation between repeated measurements
[9]	2019	England and Wales	D	8324 dwellings with at least 3 EPCs for RF 9772 dwellings with at least 3 EPCs at the postcode for RF	2008-2016	EER, inspection & lodgement dates, physical properties such as built form, floor area, flat floor level, fabric description, glazing type	EPCs for the same properties and inconsistent data across properties with same postcode.	1,2	ML (Random Forest) to identify anomalies

EPCs have been examined from three angles in academic literature: their accuracy and reliability, their function as policy instrument, and lastly the value of the input data used for alternative applications [3]. To reliably use EPCs for the latter objectives, it is essential to understand the accuracy of the data as EPC records commonly contain errors.

2.1 Validating EPC Data

Studies using EPCs for policy and alternative applications often incorporate validation methods, though these vary in depth and scope. For policy-oriented applications, such as enforcing energy efficiency standards or guiding retrofit programmes, researchers commonly validate EPC data by comparing it with independent energy audits or alternative building datasets. For instance, Jenkins et al. [5] highlighted inconsistencies in EPC ratings through cross-validation with third-party assessments, emphasising the need for greater accuracy to support policies like the MEES.

Pasichnyi et al. [10] adapted the six-level validation framework by Simon [11] to EPC data validation, ranging from basic structural checks (Level 0) to cross-referencing with external datasets (Level 5):

- Level 0, Data Structure Validation: Evaluates the format and structure of the EPC dataset, checking variable types and overall organisation but not the data itself.
- Level 1, Intra-Dataset Validation: Conducts variable consistency checks between EPCs by applying physical and numerical validation procedures.
- Level 2, Longitudinal Dataset Validation: Tracks changes over time by comparing different versions of the same dataset, therefore it is referred to as inter-dataset checks. Both EPC corrections (revision checks) and updates (time-series checks) are level 2 validation.
- Level 3, Mirror Validation: Compares the EPC dataset against alternative data sources, such as data from independent energy audits, alternative energy models, or smart meter data, to identify discrepancies. Mirror checks are inter-source intra-dataset checks.
- Level 4, Within Data Collector Validation: Cross-references the EPC dataset with additional datasets from the same data collector, such as government records or organisational databases.
- Level 5, Between Data Collectors Validation: Integrates external datasets from other organisations or regions, such as geographical or taxation data.

The source of data used for validation separates these validation levels from each other. Besides level 0 that looks at data structure, 1 and 2 validation levels are based on the same source of data (e.g. EPC dataset in isolation). In level 3 to level 5, auxiliary data sources are used. Simon [11] states that the higher levels are more challenging due to dataset accessibility, but do not necessarily mean increased accuracy or reliability. Most existing EPC accuracy studies, however, concentrate on Levels 0–3, which involve consistency checks within datasets or across repeated assessments of the same property [3,5,8,9].

Most EPC auditing studies assess accuracy by comparing inputs from repeated EPCs for the same property (level 2). In other words, they are consistency checks. This is a useful approach to identify anomalies when at least 2 EPCs are available for comparison. However, it does not determine which EPC is correct or improve the data. Additionally, many properties have only one EPC record. For example, this limited the study of Yuan and Choudhary [3] to 4% of the target building type.

Hardy and Glew [9] analysed EPCs for the same properties (level 2) and reported significant errors in EPC data, estimating that 27% of records contained at least one issue, often due to variations in assessors' interpretations of building features. The errors include duplicate records differing only by dates, unlikely changes in building features such as wall or roof types over time, and anomalies in energy efficiency products (e.g. reduced insulation or glazing performance). Many errors result from differences in opinion among EPC assessors regarding building features such as age band, floors, walls, and systems or difficulties in assessing the location of flats or maisonettes within a building or the surrounding conditions [3,9]. In interviews with 20 assessors, Gledhill et al. [6] discovered that in some cases, these variations could be intentional: "*EPCs for sale and rental market are manipulated sometimes to make a property more attractive by the estate agent and the EPC isn't as accurate as it should be*".

Errors in EPC building characteristics vary across different variables. Built form classification, particularly for end-terrace or semi-detached properties, is the most inconsistent [9]. Errors in floor and wall types were observed in 10-15% of EPCs, while wall insulation and glazing performance errors occurred in around 5% [9]. Another study found that roof efficiency and total floor area exhibited the least consistency across samples. In contrast, heating system and wall efficiency data were the most consistent [5,6]. Hardy and Glew [9] found that flats and maisonettes present more errors, often due to difficulties in assessing their location within a building or the surrounding conditions, and because the built form of flats and maisonettes is not used in SAP calculations, the assessor might not pay attention to the built form. In another study, Crawley et al. [8] found that the magnitude of error decreases as the building performance increases, whereby G-rated dwellings have more errors than A-rated ones.

Quantifying the magnitude of these errors and their impact on EER is a subject of debate. Jenkins et al.'s EPC 'Mystery Shopper' study [5] quantified the effect of these errors on EER through a level 3 mirror validation against known building info. At least four assessors evaluated the same 29 volunteer households across England and Wales, leading to an average variation of 11 points per dwelling, with one property differing by over 30 points. The results are similar to a Swedish domestic building by Hårsman et al. [12] who identified 20% variation across assessors and Crawley et al. [8] who found that the size of error could even result in wrong EPC rating. Hardy and Glew [9] concluded that an individual mistake on an EPC assessment is unlikely to make a considerable difference in the EER. In other words, identifying a single error does not necessitate removing an EPC from the analysis, although this is the most common pre-processing approach as described in section 2.2.

Higher-level validation approaches (Levels 3 to 5) leverage auxiliary datasets, such as geographic information systems (GIS), property registers, and socioeconomic data, to cross-check EPC data against other datasets. While these methods provide a more comprehensive perspective on data quality, their application is limited by challenges in accessing and integrating diverse datasets.

There is a wide range of place-based data from the Census, mostly published at, at least, one of the following geographies that were primarily designed for the publication of the Census statistics. Output Areas (OAs) are the lowest level of geographical areas. Lower Layer Super Output Areas (LSOAs) group OAs, usually four or five, containing between 400 and 1,200 households and have a resident population between 1,000 and 3,000 people. LSOAs are nested within the boundaries of Middle Layer Super Output Areas (MSOAs), which are nested within the boundaries of Local Authority Districts (LADs). These geographical areas are homogenous in terms of population size. So, one can compare like-for-like when looking at changes over time and when comparing different areas [13].

Building energy efficiency varies locally [14], with low-income households bearing higher costs of low-carbon policies in the UK due to a higher percentage of their income on energy bills [15]. EPC data linked with socioeconomic information can help identify areas at higher risk of fuel poverty [16]. Socioeconomic data at LSOA cannot be used for quantification but can be used for comparison across small areas, changes over years, and the domains/types of deprivation, particularly for financial aid and policy. For example, Yuan and Choudhary [3] demonstrated how combining EPC data with domestic gas consumption and social factors can inform a more locally tailored design of residential heating policies. No study has validated these levels to date; this study addresses that gap.

Studies focused on building energy retrofitting often use EPC recommendations and auxiliary data like GIS and property registers. Gupta and Gregg [17] designed retrofit packages for UK domestic buildings based on EPC carbon emissions and used a city-level energy mapping model to identify suitable neighbourhoods and houses for targeted retrofits, promoting bulk installations to reduce costs.

In alternative applications of EPCs, such as assessing retrofit potential, analysing fuel poverty, or evaluating the impact on real estate market value, validation often involves linking EPC data with auxiliary datasets. These include socioeconomic data, property transaction records, and GIS-based spatial analyses. For example, studies have combined EPC ratings with the Index of Multiple Deprivation (IMD) to identify areas of high energy inefficiency or fuel poverty risk [3,14]. However, these studies frequently lack systematic validation of the underlying EPC data itself, relying primarily on the end-use metrics rather than scrutinising the quality of inputs. This gap underscores the importance of integrating rigorous validation frameworks into studies leveraging EPC data for broader applications.

2.2 Improving EPC Data

Following data validation, the identified errors are either removed or corrected through substitution or imputation during the data editing stage [11]. Yet, most studies do not correct the EPC data but instead remove without quantification of the acceptable level of error in EPC data. Removing EPCs or datapoints from evaluations can introduce a bias into the results. Besides, in studies that aim to utilise EPC values for alternative applications, removing is not an option as it is necessary to determine the most accurate estimates.

One of the applications of EPCs studies their impact on property sales and rental prices [18–20]. These studies associate physical, locational, and neighbourhood characteristics to the property value. Fuerst et al. examined if there is an economic case for energy efficient dwellings in rental and purchase price. They found that 37% and 17% of EPCs in Wales from 2003 to 2014 did not record the age and number of bedrooms in turn. These EPCs were excluded from their study and whether a systematic bias was presented in non-recording EPCs is not known [18].

Not all EPC variables hold the same level of importance in analyses, and the decision to include or exclude an EPC with missing data depends on which variable is missing. Similarly, when comparing repeated EPCs, not all variations represent significant inconsistencies that warrant reporting. For instance, in existing dwellings assessed using RdSAP, certain variables are assigned default values. For example, the dwelling's type and age are used to estimate window areas, while age and wall type are used to infer U-values for walls, roofs, and floors [18]. Consequently, age (or vintage class) emerges as a critical variable.

Hardy and Glew [9] showed that ML has the potential for improving EPC auditing by identifying inconsistencies across houses within the same postcode (level 1) and EPC updates for the same house (level 2). They limited the selection to postcodes and houses with at least three EPCs, with both analyses misclassifying 7% of 'correct' data as errors while identifying 62% and 80% of the errors, respectively.

3 Methodology

A comprehensive EPC validation and improvement framework was developed following a systematic review of validation approaches within the literature, publicly available databases, and domestic EPC records in England and Wales. The framework was implemented in Python and follows the flowchart shown in Figure 1 and described below.

Leeds, the second most populous area and an important economic hub in the UK, was chosen as a case-study to illustrate the approach. Domestic EPC records in Leeds were sourced from the UK's Open Data Communities platform published by the Ministry of Housing, Communities and Local Government [21].

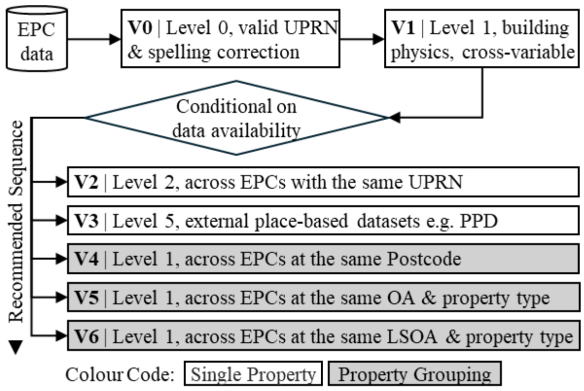

Figure 1. EPC Data Validation Engine

The six-level validation framework was applied to assess data quality of EPCs using publicly available datasets. Validation was applied based on reliability, prioritising property-specific checks over area-based analysis. This distinction is reflected in Figure 1, where property-specific validations are marked in white (V0 to V3), while validations based on property groupings are shaded in grey (V4 to V6).

Level 3 can benefit from manual user inputs or independent energy audits for mirror validation. It could include smart meter data load shape analysis and energy use disaggregation [22]. This study however is focused on automated approaches in the absence of access to building specific information. Moreover, due to unavailability of other data from the data collector, no level 4 validation is considered.

In the UK, unique property reference numbers (UPRN) are used to record anonymised public EPCs. Therefore, the level 0 validation (V0) ensures that all EPCs have a valid UPRN and no spelling inconsistencies. The dataset was then filtered to include valid EPC records within the latest 10 calendar years (2014 up to 2024), acknowledging the 10-year validity of EPCs.

Level 1, internal consistency within dataset, involves applying building physics and numerical validation procedures. V1 building physics validation is informed by leveraging insights from the literature and domain expertise. While statistical validation occurs at different tiers of geographical areas, i.e. properties within the same postcode (V4), OA (V5), and LSOA (V6). In higher geographical areas, other than properties with the same postcode, the validation is more for inconsistency checks or potential anomaly detection rather than value correction through statistical analysis and ML.

As shown in Figure 1, proceeding to the next steps beyond building physics and cross-variable checks, depends on the availability of sufficient data. For example, Level 2 (V2) compares former EPC reports of the same dwelling using UPRN, if available.

Level 5 involves cross validation with external place-based datasets, including socioeconomic data. These datasets provide contextual insights that also assist with extending the application of validated EPC data. To append auxiliary socioeconomic data, the ONS Postcode Lookup library [23] was used to retrieve OAs and LSOAs for every postcode. V3 utilises HM Land Registry Price Paid Data (PPD) [24] as an example of auxiliary property-level data. PPD reports property sales in England and Wales in every calendar year [24]. To maximise matched records for better analysis, PPD was mapped to EPCs using all possible postcode and address line combinations, ensuring inspection and price year align.

To construct datasets for anomaly detection, this study applies and enhances the approach proposed by Hardy and Glew [9], where artificial errors were introduced by modifying subsets of EPCs. UPRN and postcode were used as two identifiers to group EPCs. Only groups containing at least three EPCs were retained in V2 and V4. Within each sub-dataset, about one-third of the EPCs were randomly selected, with single building parameters replaced by the values randomly drawn from a different group. The modified EPCs were labelled as 'False', while the remaining ones were labelled as 'True'. Categorical variables were transformed based on their similarity within the same group. For example, if a postcode had three EPCs with roof types of flat, pitched, and pitched, the similarity values would be 0.33, 0.66 and 0.66, respectively. Numerical variables were transformed into percentage values relative to the mean. For example, if a UPRN has three EPCs with floor areas of 30, 40, and 50, the percentage values would be 0.75, 1, and 1.25. These features allow the Random Forest (RF) model to detect potentially erroneous records.

To further develop their approach, this study also considers two additional identifiers: (1) the combination of OAs and property types (V5) and (2) the combination of LSOAs and property types (V6). Given the larger number of properties within higher geographical areas such as OAs and LSOAs, only groups with at least ten EPCs were retained. Furthermore, they trained the model on an entire sub-dataset, which increases the risks of data leakage and model generalisation ability, causing it to perform overoptimistic conclusions due to the absence of testing datasets [25]. To mitigate this issue, the four sub-datasets were split into an 80% training set and 20% testing set, as it is empirically the best division into the training and testing sets [26]. Due to their fast computation speed and efficiency, gradient boosting classification models including CatBoost, LightGBM and XGBoost were also applied for anomaly detection, along with the RF model [27].

4 Results

391,441 EPCs were identified in the Leeds LAD (GSS code E08000035). 9,750 of these were removed due to null UPRNs (V0). 190,221 EPCs were within the considered 10 calendar years.

V1 identifies intra-variable inconsistencies in EPCs. Improving the EPC values was achieved as per the following steps depending on the variable class: numeric, e.g. floor height, or character, e.g. hot water description. Categorical variables were assessed by unique values. For example, in Leeds, there are more than 250 unique wall, floor, and roof type descriptions. In validation level 0, spelling errors that resulted in unique values were corrected. Many of these unique values refer to the same construction composition and properties, thus can be combined into a smaller list of variations. The process of standardising these variables results in creating a rule-based dictionary that associates relevant variables to cross-check values and improve the properties where needed.

The key categorical variables are the construction age bands and property types, which were harmonised. Property type can also be compared to PPD database in level 5 validation (V3) to ensure accuracy. These key variables are essential because they are the basis for assumptions about the fabric construction and systems in the building. For example, rdSAP makes assumptions about the wall composition and thermal resistance properties based on these. One example of intra-variable improvement is entries with 'Tenure' labelled as 'Unknown', whereby conditional logic was applied based on the 'Transaction type' column to infer likely tenure types. Transactions marked as private or social rentals were reassigned accordingly.

One focal example of categorical variables is when looking across fields related to the heating source and system, where gas might be mentioned as the 'main fuel', while 'mains gas flag' might be listed as N and 'main heat description' might suggest electric storage heaters. A detailed intra-variable logic was developed to review all heating and hot water related properties. Combined with other level 1 and 2 validations, these were identified and revised where possible.

Figure 2Figure **2** summarises the accuracy evaluation of various ML models (V2, V4, V5, and V6). CatBoost achieved the highest accuracy of approximately 85% in the postcode-based dataset (V4). In the UPRNs-based dataset (V2), CatBoost also outperformed other models with an accuracy of around 81%, closely followed by LightGBM, while RF showed the lowest performance. For the sub-dataset based on the combination of LSOAs or OA and property types (V5 and V6), RF showed superior performance, achieving around 84% and 85% accuracy, respectively. Overall, CatBoost and RF were the top-performing models.

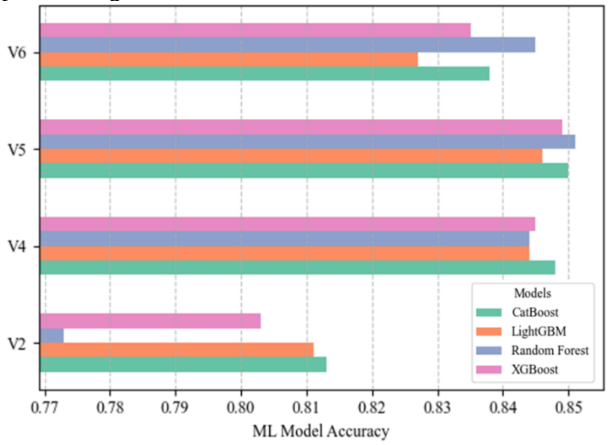

Figure 2: Accuracy comparison of ML models

Based on the accuracy results, the best-performing ML model for each validation was selected for further analysis using confusion matrices, as shown in Table 2. In V2, the model correctly classified 2622 'True' and 967 'False' EPCs, and misclassified 311 as 'False' and 517 as 'True'. While in V4, the model correctly predicted 27660 'True' and 11008 'False' records but misclassified overall 6949 EPCs.

Table 2: Confusion matrices for the best ML models

Level		False	True	Class Error
V2	False	967	517	0.34
	True	311	2622	0.10
V4	False	11008	4240	0.27
	True	2709	27660	0.08
V5	False	10740	4063	0.27
	True	2569	27103	0.08
V6	False	10532	4664	0.30
	True	2436	28150	0.07

5 Discussion

This study develops a data-driven EPC validation framework, addressing key research gaps. By integrating ML and rule-based corrections, it detects anomalies and refines data quality. Results highlight the value of auxiliary datasets, multi-level validation, and ML in anomaly detection.

5.1 Evaluation of ML Model Performance

ML models performed well across validation levels, with CatBoost and RF as top performers. In the postcode-based dataset (V4), CatBoost achieved the highest accuracy of 85%, while in the UPRNs-based dataset (V2), it led with 81% accuracy, followed by LightGBM. These results suggest location-based grouping improves predictions, likely due to spatial correlations in EPC characteristics.

For datasets combining geographical areas and property types (V5 and V6), RF excelled (85% accuracy), indicating hierarchical property aggregation enhances anomaly detection. This aligns with prior studies [9] on validation accuracy through geographical grouping. However, False Positives remain an issue, with misclassified 'True' EPCs slightly increasing in V6 (0.30) from V4 and V5 (0.27). While True class error consistently improved (0.10 in V2 to 0.07 in V6), distinguishing erroneous records remains a challenge as the dataset scales.

5.2 Implications of EPC Validation and Policy Relevance

The effectiveness of EPCs as a policy tool depends on their accuracy. The UK government's EPC reform [28], set for 2026, aims to improve reliability by introducing a fabric efficiency rating, heating system type, and an emissions classification. The proposed reforms highlight the increasing reliance on EPCs for energy policy, financial incentives, and decarbonisation planning. However, EPC accuracy issues can undermine their effectiveness in these applications [3,5,9]. The findings reinforce the need for incorporating systematic validation frameworks into EPC auditing to ensure reliability.

This study also demonstrates that auxiliary datasets, such as HM Land Registry PPD, can enhance EPC validation, offering a new avenue beyond traditional consistency checks. While previous studies linked EPCs to property pricing [18–20], this research shows that they can also be leveraged to validate EPC inputs. Future policy reforms should consider integrating external datasets into EPC validation frameworks to improve reliability at scale.

5.3 Challenges and Future Considerations

While the proposed methodology has proven effective, several limitations warrant further research. First, geographical validation (postcode, OA, LSOA) detects anomalies but does not pinpoint specific errors in EPCs. While property-level validation (V3) improves accuracy, future studies should explore hybrid approaches that integrate spatial validation with assessor verification. Mirror validation (Level 3), using smart meter data or thermal imaging, could further enhance real-world accuracy.

This framework does not account for assessor-related inconsistencies, despite evidence that assessor experience, software versions, and subjective interpretation influence EPCs [5,6]. Future research should examine assessor-driven variations and develop standardised correction mechanisms to further refine EPC datasets.

6 Conclusions

This study demonstrates that systematic validation frameworks, integrating both rule-based corrections and ML techniques, can significantly enhance EPC reliability. By applying a multi-level validation approach and leveraging auxiliary datasets, this research identifies key limitations in conventional EPC validation methods and offers a scalable, data-driven pathway for improvement. ML models achieved up to 85% accuracy, highlighting the potential of automated anomaly detection. However, False Positives and rating imbalances remain challenges, requiring further refinement. Future research should expand access to external datasets, integrate mirror validation techniques, and develop scalable, interoperable validation systems to further improve EPC accuracy and utility.

References

[1] EPBD, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), 2010. http://data.europa.eu/eli/dir/2010/31/oj/eng (accessed December 20, 2024).

[2] BEIS, Non-domestic private rented property: minimum energy efficiency standard - landlord guidance, (2023). https://www.gov.uk/guidance/non-domestic-private-rented-property-minimum-energy-efficiency-standard-landlord-guidance (accessed December 20, 2024).

[3] M. Yuan, R. Choudhary, Energy Performance Certificate renewal — An analysis of reliability of simple non-domestic buildings' EPC ratings and pragmatic improving strategies in the UK, Energy Policy 178 (2023) 113581. https://doi.org/10.1016/j.enpol.2023.113581.

[4] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Global sensitivity analysis of England's housing energy model, Journal of Building Performance Simulation 8 (2015) 283–294. https://doi.org/10.1080/19401493.2014.925505.

[5] D. Jenkins, S. Simpson, A. Peacock, Investigating the consistency and quality of EPC ratings and assessments, Energy 138 (2017) 480–489. https://doi.org/10.1016/j.energy.2017.07.105.

[6] T. Gledhill, W. Swan, R. Fitton, A practitioner study into the variability of UK domestic energy assessments, International Journal of Building Pathology and Adaptation ahead-of-print (2023). https://doi.org/10.1108/IJBPA-10-2022-0167.

[7] Scottish Government, Energy Performance Certificate (EPC) reform: consultation, 2023. https://www.gov.scot/publications/energy-performance-certificate-epc-reform-consultation/ (accessed December 24, 2024).

[8] J. Crawley, P. Biddulph, P.J. Northrop, J. Wingfield, T. Oreszczyn, C. Elwell, Quantifying the Measurement Error on England and Wales EPC Ratings, Energies 12 (2019) 3523. https://doi.org/10.3390/en12183523.

[9] A. Hardy, D. Glew, An analysis of errors in the Energy Performance certificate database, Energy Policy 129 (2019) 1168–1178. https://doi.org/10.1016/j.enpol.2019.03.022.

[10] O. Pasichnyi, J. Wallin, F. Levihn, H. Shahrokni, O. Kordas, Energy performance certificates — New opportunities for data-enabled urban energy policy instruments?, Energy Policy 127 (2019) 486–499. https://doi.org/10.1016/j.enpol.2018.11.051.

[11] A. Simon, Definition of validation levels and other related concepts, European Commission. Eurostat., 2013. https://circabc.europa.eu/sd/a/43c3c89d-8ab4-4f39-af3f-07185035ff80/ASA_AAP_154_Annex%201_Definition_of_validation_levels_etc.doc (accessed December 24, 2024).

[12] B. Hårsman, Z. Daghbashyan, P. Chaudhary, On the quality and impact of residential energy performance certificates, Energy and Buildings 133 (2016) 711–723. https://doi.org/10.1016/j.enbuild.2016.10.033.

[13] ONS, 2011 Census geographies - Office for National Statistics, (2024). https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/2011censusgeographies (accessed January 15, 2025).

[14] J. Ahlrichs, S. Wenninger, C. Wiethe, B. Häckel, Impact of socio-economic factors on local energetic retrofitting needs - A data analytics approach, Energy Policy 160 (2022) 112646. https://doi.org/10.1016/j.enpol.2021.112646.

[15] A. Owen, J. Barrett, Reducing inequality resulting from UK low-carbon policy, Climate Policy 20 (2020) 1193–1208. https://doi.org/10.1080/14693062.2020.1773754.

[16] R. Camboni, A. Corsini, R. Miniaci, P. Valbonesi, Mapping fuel poverty risk at the municipal level. A small-scale analysis of Italian Energy Performance Certificate, census and survey data, Energy Policy 155 (2021) 112324. https://doi.org/10.1016/j.enpol.2021.112324.

[17] R. Gupta, M. Gregg, Targeting and modelling urban energy retrofits using a city-scale energy mapping approach, Journal of Cleaner Production 174 (2018) 401–412. https://doi.org/10.1016/j.jclepro.2017.10.262.

[18] F. Fuerst, P. McAllister, A. Nanda, P. Wyatt, Energy performance ratings and house prices in Wales: An empirical study, Energy Policy 92 (2016) 20–33. https://doi.org/10.1016/j.enpol.2016.01.024.

[19] F. Fuerst, M.F.C. Haddad, H. Adan, Is there an economic case for energy-efficient dwellings in the UK private rental market?, Journal of Cleaner Production 245 (2020) 118642. https://doi.org/10.1016/j.jclepro.2019.118642.

[20] F. Fuerst, P. McAllister, A. Nanda, P. Wyatt, Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England, Energy Economics 48 (2015) 145–156. https://doi.org/10.1016/j.eneco.2014.12.012.

[21] Department of Levelling Up, Housing & Communities, Energy Performance of Buildings Data England and Wales, (2024). https://epc.opendatacommunities.org/domestic/search (accessed December 21, 2024).

[22] M. Taheri, L. Jacob, C. Parry, S. Mirzaie, A. Hermanowicz, I. Vrachimi: Alan, The efforts towards development of an energy efficiency upgrade platform, in: IBPSA, 2021: pp. 3300–3307. https://doi.org/10.26868/25222708.2021.30910.

[23] Postcode to OA (2021) to LSOA to MSOA to LAD (November 2024) Best Fit Lookup in the UK, (n.d.). https://open-geography-portalx-ons.hub.arcgis.com/datasets/ons::postcode-to-oa-2021-to-lsoa-to-msoa-to-lad-november-2024-best-fit-lookup-in-the-uk/about (accessed March 14, 2025).

[24] HM Land Registry, Price Paid Data, GOV.UK (2024). https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed December 24, 2024).

[25] M.A. Bouke, A. Abdullah, An empirical study of pattern leakage impact during data preprocessing on machine learning-based intrusion detection models reliability, Expert Systems with Applications 230 (2023) 120715. https://doi.org/10.1016/j.eswa.2023.120715.

[26] A. Gholamy, V. Kreinovich, O. Kosheleva, Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation, International Journal of Intelligent Technologies & Applied Statistics 11 (2018) 105–111. https://doi.org/10.6148/IJITAS.201806_11(2).0003.

[27] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif Intell Rev 54 (2021) 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

[28] DESNZ and MHCLG, Open consultation - Reforms to the Energy Performance of Buildings regime, GOV.UK (2024). https://www.gov.uk/government/consultations/reforms-to-the-energy-performance-of-buildings-regime (accessed December 24, 2024).