Annals of Operations Research
https://doi.org/10.1007/s10479-026-07187-w

ORIGINAL RESEARCH

Predicting energy poverty using household budget survey: a machine learning approach

Giuseppe Scandurra[1] · Alfonso Carfora[2] · Antonio Thomas[3] · Cecilia Camporeale[4]

Received: 10 January 2025 / Accepted: 25 March 2026

Abstract

Energy poverty (EP) is considered an urgent challenge, intensified by rising energy costs, economic inequality, and the transition toward green energy, which involves many Western countries. By referring to Italy, this study employs machine learning algorithms (MLAs) to predict and classify EP using official Household Budget Survey (HBS) data. To evaluate EP, the study compares several MLAs alongside three expenditure-based indicators proposed in three seminal articles by Hills, Faiella and Lavecchia, and Betto et al. Among these, the indicator developed by Betto et al., which accounts for regional and socioeconomic disparities, consistently outperforms the others across all MLAs, demonstrating higher accuracy, precision, and recall. This ensures a more comprehensive identification of energy-poor households. The analysis highlights the significant impact of data imbalance on model performance, emphasizing the need for techniques such as SMOTE and undersampling. The superior performance of the Betto et al. indicator underscores its potential as a benchmark for EP measurement, providing a valuable tool for policymakers to design targeted interventions, allocate resources effectively, and support a just and sustainable energy transition. The study reinforces the importance of dynamic, data-driven approaches to address EP, and calls for improved data collection to enhance prediction accuracy and policy effectiveness.

Keywords Energy poverty · Household budget survey · Random forest · Boosting · Staking

1 Introduction

Energy poverty is a multifaceted and pervasive issue that poses significant challenges in the ensuing decades (Primc & Erker, 2020; Stojilovska et al., 2022; Kashour & Jaber, 2024) due to the consideration that energy plays a central role in modern society and lifestyle, serving

✉ Giuseppe Scandurra
giuseppe.scandurra@uniparthenope.it

1 Department of Management Studies and Quantitative Methods, University of Naples Parthenope, Naples, Italy

2 Department of Law and Economics, University of Macerata, Macerata, Italy

3 Department of Business and Economics, University of Naples Parthenope, Naples, Italy

4 Technical and Strategic Support Unit of Department for Sustainability, ENEA, Roma, Italy

Published online: 13 April 2026

as a fundamental commodity that supports various activities from basic, such as heating, lighting, and cooking, to the complex ones, i.e. transport, etc.

Energy poverty (henceforth EP) represents one of the most important issues for policy-makers and researchers by virtue of the increasing share of households with difficulties in adequately warming (and cooling) their apartments (Henry et al., 2021; Zhang et al., 2022). Thus, both policymakers and researchers are urged to look at this problematic requires increasing attention due to the combined effect of the inflationary downturn that has hit the energy sector (Carfora & Scandurra, 2024), with the post-pandemic and international tensions and the necessity to accelerate the transition to cleaner and more reliable energy systems (IEA, 2024) tackling EP (Leimbach & Giannousakis, 2019; Zhao et al., 2022).

In this view, Government policies and consumer choices will have huge consequences for the future of the energy sector and the fight against climate change Despite the consensus on the urgency of taking actions to limit the impacts of climate change and the growing momentum towards a transition to clean energy, the world is still far from a trajectory aligned with achieving the goal of the Paris Agreement, leading instead to an *energy poverty trap* (Hussain et al., 2023).

However, EP is a phenomenon that affects all countries in the world, both low-income (Al Ketz et al., 2024; Babayomi et al., 2022; Banerjee et al., 2021) and higher-income countries (Hasheminasab et al., 2023; Katoch et al., 2024; Lu & Ren, 2024; Thomson et al., 2017), albeit for different reasons and in various ways.

Across the European Union (EU), EP is a critical issue affecting over 41 million people, representing 9.3% of the population (Widuto, 2023). Thus, it has gained considerable attention, spurring extensive academic research and policy initiatives (Castaño-Rosa et al., 2020; Bouzarovski, 2021). The EU has recognized the importance of addressing EP, prompting regulatory actions such as Directive 2009/72/EC and Regulation 2018/1999, which mandate member states to protect vulnerable customers and integrate EP considerations into their national energy and climate plans.

Even so, addressing EP remains a complex task due to the varied reasons it originates, the dimensions to interpret the phenomenon, and its manifestations. Moreover, the lack of a universally accepted framework for defining and understanding EP, and comparable data to measure it, further complicates mitigation efforts (Roberts et al., 2015; Faiella & Lavecchia, 2021). For instance, even today, few European countries—specifically France, Ireland, the United Kingdom, Cyprus, and Slovakia—have adopted an official definition of EP (e.g., Dobbins et al., 2019; Thomson and Bouzarovski, 2019; Al Ketz et al., 2024b). Often qualified as the inability of households to afford adequate energy services for a healthy and comfortable living environment (e.g., Boardman, 1991), EP also encompasses both social and material aspects.

EP is a particularly acute issue in Italy, where high energy costs, low household incomes, inefficient housing stock, and significant regional economic disparities contribute to its persistence and growth. According to the European Energy Poverty Advisory Hub (EPAH), Italy is experiencing a moderate but steadily worsening level of EP, with conditions gradually deteriorating since 2017.[1] Data from the Italian Energy Poverty Observatory indicate that in 2021, 8.5% of households were affected. A key element of Italy's strategy to address EP involves offering fiscal incentives and tax exemptions to support energy-efficient housing renovations. These facilitations are designed to incentivize homeowners to undertake energy-efficient renovations, thereby reducing energy consumption and alleviating the financial burden on households. The effectiveness of these renovation grants in reducing EP (Santangelo et al.,

[1] https://energy-poverty.ec.europa.eu/epah-indicators.

2019) is a subject of great interest, as it has implications for household energy consumption and cost but also for broader environmental and economic goals (Kyprianou et al., 2019).

However, despite the measures adopted to promote greater building efficiency and, consequently, to tackle EP, there is a need for the correct identification of households that are energy-vulnerable (e.g., Dogan et al., 2021; Middlemiss, 2022). As noted in the existing literature (e.g., Pye et al., 2015; Tovar, 2021), interventions to mitigate EP and address income poverty may require different approaches. Therefore, it is necessary to (i) recognize energy-poor families, (ii) identify the areas where interventions are needed, and (iii) isolate those factors that most significantly aggravate energy deprivation (Castaño-Rosa et al., 2020; Dubois, 2012). To effectively support households unable to afford their energy needs, any EP mitigation strategy should incorporate a framework that allows policymakers to identify such households using the available data, which are currently insufficiently detailed and have incomplete coverage (Sareen et al., 2020). This approach would allow the precise targeting of mitigation measures, avoiding reliance on self-reported household energy deprivation.

Considering this knowledge gap, this study compares three machine learning algorithms (MLAs) to predict and target Italian households in EP using high-dimensional data from the Household Budget Survey (HBS) provided by the Italian National Statistical Institute (ISTAT). The HBS data can be used to calculate the expenditure-based EP indicators, providing a comprehensive picture of EP in the country.

As is well known, various approaches have been presented in the literature for measuring EP. They are based on self-reported indicators (consensual approach), direct indicators of thermal comfort, and expenditure-based methods. The classes of indicators proposed for measuring EP have limitations and strengths, mostly centered on data availability, measurement difficulties, and/or the subjectivity of some measures. A recently published paper by Al Ketz et al. (2024) summarizes their strengths and weaknesses.

Building on previous research (e.g., Abbas et al., 2022; Al Kez et al., 2024a, 2024b; Spandagos et al., 2023), this study compares three MLAs for classifying households experiencing energy poverty (EP), using expenditure-based indicators developed by Hills, (2011), Faiella and Lavecchia (2014), and Betto et al. (2020). These indicators serve as target variables for the MLAs, which are proposed as benchmark models for identifying energy-poor households. The capacity of MLAs to capture non-linear relationships and complex interactions among predictors makes them particularly well-suited for predicting EP. Using a high-dimensional dataset from HBS, these models can uncover hidden patterns and trends that traditional methods may overlook. Moreover, using data from HBS allows for overcoming one of the main limitations related to data availability.

MLAs are increasingly used to predict EP by leveraging indicators from HBS. These algorithms enhance the accuracy of identifying at-risk households and support the development of more targeted policies and interventions. Additionally, MLAs can highlight priority areas for policymakers and help address challenges related to data scarcity. Moreover, using a current and readily accessible data source, such as the mentioned HBS, is crucial to ensure that adjustments and interventions can be implemented without significant delays. This timeliness is vital, as it helps prevent any negative impact on the quality of life for the affected households. By enabling prompt and data-driven decision-making, these algorithms contribute to a more efficient and effective approach to addressing EP, ensuring that resources are directed where they are most needed.

Furthermore, this study sheds light on fiscal bonuses for housing renovations, helping to assess how they contribute to predicting and alleviating EP. The hypothesis is that these budgetary measures not only help reduce energy consumption but also improve the financial resilience of households, thereby reducing their vulnerability to EP.

This study adds to the expanding body of research on EP by applying MLAs to the Italian context, an area that has thus far been underexplored in data-driven EP studies. Through the use of ensemble models, the analysis achieves not only strong predictive performance but also identifies the most influential factors contributing to EP, offering valuable insights for targeted policy design. The distinct contribution of this work lies in its combination of predictive analytics with model interpretability, facilitating a more nuanced understanding of the complex and multidimensional determinants of energy poverty.

By integrating quantitative data analysis with policy evaluation, this study aims to provide a robust understanding of the dynamics of EP in Italy and the role of housing renovation grants in mitigating this issue. The findings will offer valuable insights for policymakers, enabling them to design more effective interventions to support vulnerable populations and promote a sustainable energy transition.

After this introduction, Sect. 2 discusses the EP measurement, Sect. 3 discusses the method adopted in this study, and Sect. 4 reports the results of the investigations. Section 5 outlines the discussion, while the conclusion and policy implications are in the last Section.

2 Measuring energy poverty

EP is usually measured by applying various indicators that capture the inability of households to afford adequate energy services. To ensure conceptual clarity and methodological applicability, the proposed metrics for EP assessment can be categorized into three primary approaches:

(i) Direct Indicators of thermal comfort conditions, such as indoor temperature and humidity levels; (ii) Subjective or self-reported indicators approach, commonly associated with the consensual approach, which capture individuals' perceived ability to maintain adequate warmth or meet essential energy needs; and (iii) Expenditure-based approach, which evaluate the proportion of household income allocated to energy consumption relative to established affordability thresholds.

(i) *Direct Indicators of Thermal Comfort Conditions* involve assessing home conditions to determine whether they meet acceptable warmth and energy efficiency standards. These indicators provide objective data on the living conditions of households. This approach includes (see, e.g., Barnard et al., 2018; Bramley et al., 2017; Kahouli & Okushima, 2021) Indoor Temperature Measurements to determine whether they meet recommended levels for health and comfort and the Energy Efficiency of Dwellings, evaluating the energy efficiency of homes, considering factors like insulation, heating systems, and overall building performance (Billio et al., 2024; Palma & Gouveia, 2022).

(ii) *Self-Reported (consensual-based) Indicators* approach involves collecting information directly from households about their experiences and perceptions of EP (Herrero, 2017). These measures capture subjective aspects that expenditure-based measures might miss. Among the various indicators usually employed to classify energy poor's households, the Inability to Keep Home Warm is undoubtedly one of the most used to measure EP in EU countries (Castaño-Rosa et al., 2020; Dogan et al., 2021) and EU statistics on income and living conditions (EU-SILC) provide the indicators to measure EP.

(iii) E*xpenditure-based* approach is one of the most widespread for assessing EP. It primarily focuses on the relationship between household income and energy expenditure. In doing so, it relies on quantitative data. This approach can be split into three methodologies:

(1) *Energy expenditure share* (2 M) calculates the percentage of a household's income spent on energy costs. This approach classifies households as energy poor if they allocate more than a specified proportion of their income, commonly set at 10%, to cover energy-related expenses. Known as the expenditure-based threshold method, this metric assumes that spending above this benchmark indicates a disproportionate financial burden, potentially compromising other essential needs.
(2) *Low Income High Costs* (LIHC) indicator identifies households as energy poor when they simultaneously experience above-average required energy costs and have a residual income, after accounting for these costs, that falls below the nationally defined poverty line. This dual-criterion approach captures both the burden of high energy expenditures and the insufficiency of disposable income, offering a more nuanced perspective than single-threshold methods. By grounding the poverty line in national standards, the LIHC framework allows for contextual sensitivity to local economic conditions, while maintaining comparability across different regions and demographic groups
(3) *Residual energy expenditure* compares a household's actual energy spending to the national or regional average (or median). Under this methodology, households that spend significantly more than the average (or median) may be classified as experiencing EP.

Leading scholars have applied expenditure-based indicators to study EP (Betto et al., 2020; Faiella & Lavecchia, 2015; Hills, 2011). These indicators have proven helpful in analyzing EP in Italy (e.g., Bardazzi et al., 2024) and other EU countries (e.g., Valeria, 2024).

To provide a more comprehensive approach to measure EP, scholars have proposed multidimensional indicators to capture various aspects of vulnerability, from economic constraints to energy access and consumption patterns (see, e.g., Nussbaumer et al., 2013; Scandurra et al., 2024; Thomson & Snell, 2013). One of the first indicators was the *Multidimensional Energy Poverty Index* (MEPI) (Nussbaumer et al., 2012, 2013), which combines various indicators such as access to energy, energy affordability, and energy efficiency of the household to provide a holistic measure of EP. Thomson and Snell (2013) highlighted the utility of indicators derived from the *European Survey on Income and Living Conditions* (EU-SILC) in developing composite measures. They focused on creating a weighted average of three indicators to offer a more comprehensive assessment.

Kashour and Jaber (2024) introduced a novel composite index based on EU-SILC consensual indicators. Their index serves a dual purpose: ranking the EU Member States according to their levels of EP and identifying the key drivers behind this issue. Similarly, Carfora and Scandurra (2024) employed the EU-SILC indicator of "inability to keep the home adequately warm" to forecast the medium-term evolution of EP. Their study considered the inflationary pressures arising from the COVID-19 pandemic and the ongoing Russian-Ukrainian crisis as major influences on this trend.

Each class of indicators proposed for measuring EP includes strengths and limitations, often related to data availability, measurement challenges, or the subjectivity of specific measures (e.g., Al Ketz et al., 2024). As a result, these methods are seldom employed in the literature. Each approach offers unique insights into EP, and the choice of method should consider the specific context, available data, and measurement objectives.

Conversely, considerable attention has been given to indicators based on the perception of energy discomfort, particularly in Europe, where such data can be directly sourced from the EU-SILC survey. This allows for cross-country comparisons. However, this approach is not without limitations, mainly due to the self-reported nature of the data. The accuracy of responses can vary depending on how respondents interpret the survey questions, introducing a certain degree of subjectivity. Although the self-reported method can offer valuable insights into the prevalence of EP, estimates based on subjective questions have to be interpreted cautiously, as they may not always be reliable (Al Ketz et al., 2024).

The expenditure-based approach links household income with energy spending and can be quantified using thresholds established in the literature. However, while this approach is more objective, it presents challenges when attempting cross-country comparisons (Rademaekers et al., 2016; Thomson et al., 2017). Despite these limitations, particularly those related to income or expenditure thresholds, expenditure-based indicators reduce subjectivity and enable more objective household comparisons (Panão, 2021). Additionally, the HBS survey conducted by the mentioned ISTAT provides valuable data for measuring EP, as it collects variables and descriptors that can be used to achieve the goals of this study.

For all these reasons, this paper focuses on the class of expenditure-based indicators.

3 Data and method

This study focuses on the issue of EP, utilizing the data available from the 2022 Household Budget Survey (HBS) conducted by ISTAT. The HBS is a comprehensive statistical study designed to collect data on Italian household consumption patterns and other socio-demographic characteristics. As a key source for assessing household economic conditions, the survey offers critical insights that inform both policy-making and commercial decision-making processes.

The 2022 edition of the HBS involved a representative sample of 28,416 households, stratified to reflect the demographic and geographic diversity of the population at the NUTS2 (regional) level. Data collection employed a mixed-method approach, combining direct interviews with self-administered questionnaires. Respondents recorded their daily expenditures and self-consumption practices, allowing detailed categorization of household expenditure in various domains, including housing, water, electricity, gas, and other fuels.

The resulting dataset facilitates the examination of consumption trends over time, the identification of socio-economic drivers of expenditure patterns, and comparisons across income groups and demographic segments. Notably, ISTAT makes both aggregated results and anonymized microdata publicly available through its official website.[2] Drawing on this dataset, it is possible to calculate expenditure-based indicators of EP as proposed in the literature and to construct a robust set of predictor variables for in-depth analytical work.

3.1 Energy poverty in Italy using expenditure-based indicators

In alignment with the dataset used in this paper, which lacks consensual measures of EP, we use a subset of three expenditure-based indicators as outcome variables. They are (i) the Low-Income, High-Costs (LIHC) indicator, proposed by Hills (2011), (ii) the Faiella and Lavecchia (2015) indicator, and (iii) the Betto et al. (2020) indicator.

The first indicator, proposed by Hills (2011), is a composite measure that considers high energy costs with residual disposable income below the poverty level (Low-Income, High-Costs—LIHC).

[2] https://www.istat.it/en/microdata/household-budget-survey/.

This indicator, hereafter labeled as *ENPOV1*, considers household budget vulnerability to falling below the poverty line:

$$ENPOV1 = \frac{1}{n}\sum_{i=1}^{n} w_i\left(I\left(s_{ie}^{eq} > P50_t\left(s_{ie}^{eq}\right)\right) * I\left(\left(y_i^{eq} - s_{ie}^{eq}\right) < y_J^*\right)\right) \tag{1}$$

where i is the household index (i = 1, …, n), w_i are the household weights of the sample, s_{ie}^{eq} is the equivalent energy expenditure of the *ith*-household, $P50_t\left(s_{ie}^{eq}\right)$ is the *50th* percentile of the overall distribution of equivalent energy expenditure, and y_J^* is the income poverty threshold.

The second indicator is an adaptation of LIHC proposed by Faiella and Lavecchia (2015), hereafter labeled as *ENPOV2*. To identify 'hidden energy-poor households', those that cannot be identified with *ENPOV1*, it includes vulnerable households with low equivalent expenditure[3] and no heating expenditure. It is:

$$ENPOV2 = \frac{1}{n}\sum_{i=1}^{n} w_i\left(I\left(\frac{s_{ie}^{eq}}{s_i^{eq}} > 2 \times \frac{\sum_{i=1}^{n} s_{ie}^{eq}}{\sum_{i=1}^{n} s_i^{eq}}\right)\right)$$
$$* I\left((s_i - s_{ie}) < s_J^*\right) \cup \left(I\left(s_i^r = 0\right) * I\left(s_{ie}^{eq} < P50_t\left(s_i^{eq}\right)\right)\right) \tag{2}$$

Similarly to Eq. 1, i is the household index ($i = 1$, …, n); w_i are the households' weights of the sample; s_{ie}^{eq} is the equivalent energy expenditure of the *ith*-household; s_i^{eq} measures the equivalent total expenditure of the *ith*-household. Subsequently, s_{ie} measures the energy expenditure of the *ith*-household; s_i measures the total expenditure of the *ith*-household and s_J^* is the expenditure poverty threshold. $\frac{\sum_{i=1}^{n} s_{ie}^{eq}}{\sum_{i=1}^{n} s_i^{eq}}$ measures the average sample incidence of the equivalent energy expenditure, while $P50_t\left(s_i^{eq}\right)$ is the *50th* percentile of the distribution of the overall equivalent expenditure. Subsequently, s_{ie} measures the absolute (not adjusted for the equivalent scale) energy expenditure of the *ith*-household; s_i measures the absolute overall expenditure of the *ith*-household and s_J^* is the expenditure poverty thresholds. Finally, s_i^r measures the heating expenditure of the *ith*-household.

The third indicator*, ENPOV3, is* the regional EP indicator developed by Betto et al. (2020). In more detail:

$$ENPOV3 = \frac{1}{n}\sum_{i=1}^{n} w_i\left(\left(\left(s_{ie}^{eq}\right)_k < \left(\left(\frac{1}{n}\right)\sum_{i=0}^{n} s_{ie}^{eq}\right)_k\right) \cap (RP_i = 1)\right) \tag{3}$$

where RP_{i} is the relative poverty condition of the *ith*-household, and n is the number of households living in the same *k*-region (k = 1,…, 20) of the i^{th}-household. In this indicator, the threshold value is $\left(\left(\frac{1}{n}\right)\sum_{i=0}^{n} s_{ie}^{eq}\right)_k$ and represents the mean value of the energy expenditures of households living in the same region (k) of the *ith*-household. Relative poverty (RP) is a categorical dummy variable provided by ISTAT (1 if the *ith*-household is in a relative poverty condition, 0 otherwise).

[3] Equivalent expenditure and income make the expenditure and income levels of households of different sizes and compositions comparable. They are calculated by dividing income and expenditure by an appropriate equivalence scale, which makes it possible to take into account the effect of differently composed households. The equivalence scale used in this paper (called "modified OECD" and also used at the European level) has been provided by ISTAT and assigns a value of 1 to the household head, of 0.5 to each additional adult member, and of 0.3 to each child.

3.2 Methods

Applying MLAs to classification problems is widespread due to their robust performance, as highlighted in specialized literature (e.g., Chowdhury, 2024; López-García et al., 2023). Moreover, in recent years, MLAs have emerged as powerful tools for predicting EP due to their ability to analyze high-dimensional datasets (Mukelabai et al., 2023; Garibay et al., 2023). Their increasing popularity can be attributed to their proficiency in identifying patterns and relationships within high-dimensional datasets, facilitating more accurate predictions.

One of the key strengths of MLAs lies in their ability to integrate heterogeneous data sources, such as satellite imagery, household surveys, and socioeconomic indicators, to enhance predictive accuracy. Deep learning methods represent a more advanced subset of machine learning techniques. For example, Convolutional Neural Networks (CNNs), are particularly effective in analyzing visual data, such as satellite images, thanks to convolutional layers that detect patterns and image features. These deep learning models are particularly suitable for capturing interactions between multiple predictors, leading to enhanced predictive accuracy. For instance, a study by Xu et al. (2021) combined several MLAs with nighttime light remote sensing data to identify poverty-stricken areas and predict poverty levels. Puttanapong et al. (2022) demonstrated how combining satellite imagery with household survey data can significantly enhance the predictive power of MLAs in EP-related research**.** Similarly, Jean et al. (2016) demonstrated that CNNs trained on satellite imagery and survey data could explain up to 75% of local economic variation across five African countries, offering a scalable proxy for energy-related deprivation.

In Europe, Papada (2022) applied a multilayer perceptron (MLP) neural network to predict objective EP indicators, such as the 10% actual/required expenditure thresholds and the Compression of Energy Needs, based on subjective household responses. The model achieved moderate predictive accuracy (56–58%), highlighting the potential of Artificial Neural Networks (ANNs) to bridge the gap between perceived and measured EP. A broader review by López-Vargas (2022) confirmed that ANNs and decision trees are the most frequently used AI techniques in EP research, particularly for characterizing low-income households, high energy costs, and inefficient housing. Collectively, these studies illustrate the growing relevance of neural networks and MLAs in EP research, offering robust tools for early detection, policy targeting, and the design of context-sensitive interventions.

However, the decision to use the three algorithms instead of neural networks is due to the fact that the ANNs, being "black boxes", do not provide insights into the most influential variables (see, e.g., Ljung et al., 2020, Sun and Braatz, 2021).

Identifying the key variables that influence EP represents a significant analytical challenge due to the multidimensional nature of the phenomenon and the wide range of contributing factors, including income levels, housing quality, geographic location and climatic conditions. This complexity often limits the effectiveness of traditional analytical methods, which tend to rely on restrictive assumptions and are less able to capture non-linear relationships and intricate interactions between variables.

Conversely, classical MLAs such as Random Forest and XGBoost offer a more flexible and robust analytical framework. These models are particularly well-suited to high-dimensional, heterogeneous datasets and can effectively address issues such as class imbalance, a frequent concern in EP-related research, through strategies like class weighting and resampling. Furthermore, their capacity to estimate feature importance without relying on strong parametric assumptions makes them powerful tools for identifying the most influential predictors of EP. Hence, MLAs provide valuable support for evidence-based policy design by facilitating the development of targeted and context-sensitive interventions.

To classify the families that suffer a condition of EP, we compare various MLAs that can help policymakers to consider the effects of economic measures on the distribution of EP in Italian regions using the data provided by ISTAT with the HBS survey. To reach this aim, we use three well-known methods: Random Forest, Boosting, and stacking algorithms.

Instead of other classification methods (such as Support Vector Machine), the selected MLAs are preferred for their scalability, interpretability, and ability to handle high-dimensional data without excessive preprocessing. Support Vector Machine is undoubtedly a powerful choice for specific scenarios, but its limitations in scalability and interpretability make ensemble methods more attractive (e.g., Gatera et al., 2023).

3.2.1 Random forest

The Random Forest model, introduced by Breiman (2001), is an ensemble technique that builds multiple decision trees during training and aggregates their predictions for more accurate and stable results (e.g., Barsi and Arif, 2021; Chang et al., 2024). It creates multiple decision trees (usually hundreds) by training each one on a random subset of the data. Each tree is trained on a bootstrapped sample of the original dataset. At each split in the tree, Random Forest selects a random subset of features to decide the best split. This ensures that the trees are less correlated with each other and that diverse models are built. The final prediction is made by “majority voting”. The class that receives the most votes from the individual trees is chosen as the output (see, e.g., Casarin et al., 2021).

Growing multiple deep decision trees reduces the overfitting that individual decision trees can suffer from. Moreover, since the trees are built using different samples and features, the overall variance of the model is reduced, making Random Forest more robust to data heterogeneity.

3.2.2 Boosting

Boosting is another ensemble technique that builds models sequentially, with each new model attempting to correct the errors of its predecessors (e.g., Bühlmann & Hothorn, 2007). Unlike Random Forest, which builds trees independently, Boosting focuses on building trees that "boost" the performance of the previous models. This algorithm trains the model sequentially, and each new tree tries to correct the errors made by the previously trained models, assigning a weight based on the accuracy of each estimated model. Misclassified examples are given more weight in subsequent models, forcing the algorithm to focus on hard-to-classify cases. The weighted voting for classification represents the output.

Until now, literature has proposed various boosting algorithms, e.g., Adaptive Boosting (Chengsheng et al., 2017; Wu et al., 2020), Gradient Boosting (e.g., Brownlee, 2020; Krauss et al., 2017) and its optimized version, the XGBoosting (Chen & Guestrin, 2016; Mo et al., 2019). In particular, XGBoosting minimizes residual errors and adds features like regularization, faster training through tree-pruning techniques, and more effective handling of categorical variables.

3.2.3 Stacked generalization

Stacked Generalization (or Stacking) is an advanced ensemble learning technique that combines multiple machine learning models (often referred to as "base learners" or "level-0 models") to produce a stronger overall model (see, e.g., Ren et al., 2016; Wolpert, 1992).

Table 1 Expenditure-based energy poverty indicators

Variable	Poverty definition	Source
ENPOV1	Expenditure-based – Low Income, High Cost	Hills (2011)
ENPOV2	Expenditure-based – Modified Low Income, High Cost	Faiella and Lavecchia (2015)
ENPOV3	Energy expenditure is below the energy expenditure threshold and, at the same time, they are in relative poverty	Betto et al. (2020)

Unlike other ensemble methods like Random Forest or Boosting, which aggregate models in more straightforward ways, Stacking uses a "meta-model" (also called a "meta-learner" or "level-1 model") to learn how to combine the predictions of the base models best (Wang et al., 2020).

In a nutshell, staking trains various models on the same dataset, using multiple algorithms (in our analysis, decision trees, logistic regression, and boosting). The key idea in stacking generalization is that the base learners should be diverse, meaning they should make different kinds of errors so the meta-learner can combine them to improve performance. The base learners' predictions represent the input for the meta-learner. The meta-learner is typically a simple algorithm like linear regression or logistic regression (for classification), but it could also be a more complex model. The goal is to weigh the predictions from the base models to produce the most accurate final prediction (Table 1).

4 Results

This paper aimed to compare MLAs to classify households experiencing EP in Italy, starting from HBS data provided by the ISTAT, which offers detailed information on the income and expenditure of Italian households. We classify families into energy-poor and non-energy-poor groups based on the three expenditure-based indicators proposed by the quoted Hills (2011), Faiella and Lavecchia (2015), and Betto et al. (2020), which offer different perspectives on measuring EP.

Our algorithm leverages these indicators to classify households and assess the distribution of EP across different segments of the population.

Table 2 summarizes the distribution of Italian households based on these three indicators, highlighting the proportion of energy-poor and non-energy-poor households.

Table 2 Distribution of energy-poor households using the three expenditure-based indicators

Indicators	Non-energy poor	Energy poor	Total	Non-energy poor (%)	Energy poor (%)	Total (%)
ENPOV1	27,262	1154	28,416	95.94%	4.06%	100.00
ENPOV2	26,238	2178	28,416	92.34%	7.66%	100.00
ENPOV3	26,101	2315	28,416	91.85%	8.15%	100.00

The results indicate varying levels of EP depending on the indicator used. Hills's (ENPOV1) indicator identifies 4.06% of households as energy-poor, while Faiella and Lavecchia's approach (ENPOV2) yields a higher estimate of 7.66%. The indicator by Betto et al. (ENPOV3) includes additional socioeconomic factors and identifies a slightly higher proportion of energy-poor households at 8.15%.

These variations suggest that the choice of indicator significantly impacts the measured prevalence of EP. Hills' method, which focuses on income and energy expenditure thresholds, may be more sensitive to households with high energy costs. In contrast, Faiella and Lavecchia's method, centered on affordability, likely identifies a more conservative subset of energy-poor households. Betto et al.'s approach, which takes a broader perspective, seems to capture a more comprehensive picture of EP in Italy.

Focusing on data in Table 2, we observe that the number of energy-poor households is underrepresented compared to the number of non-energy-poor households across all three indicators (ENPOV1, ENPOV2, and ENPOV3). The proportion of energy-poor households ranges from around 4% to 8%. In comparison, non-energy-poor households represent more than 90% of the total population, highlighting the presence of imbalanced data that can cause various challenges for machine learning models, especially in classification tasks (e.g., López et al., 2013; Thabtah et al., 2020).

When applying MLAs to such imbalanced datasets, the model may become biased toward the majority class (non-energy-poor households), resulting in poor performance in identifying the minority class (energy-poor households).

Many MLAs tend to predict the majority class more frequently because it minimizes the overall error rate. As a result, the model may not identify energy-poor households accurately. Even though the overall accuracy might seem high (because the model is correct most of the time for non-energy-poor households), the recall and precision for the minority class (energy-poor households) could be very low.

To overcome the issue of class imbalance, we can use several strategies when applying machine learning algorithms, such as resampling techniques and class weighting.[4]

As resampling, we undersample the majority class by reducing the number of non-energy-poor samples to match the energy-poor samples, and the Synthetic Minority Over-sampling Technique (SMOTE) that creates synthetic examples of the minority class by interpolating between existing samples, increasing the diversity of the minority class.

Moreover, many MLAs allow class weights to be assigned. By assigning a higher weight to the minority class (energy-poor households), the model becomes more sensitive to correctly identifying those cases. This is particularly useful for algorithms like Random Forests or XGBoost that can automatically adjust for class imbalance with proper configuration.

While SMOTE is a popular method for addressing class imbalance, it should be applied with caution and is generally not recommended in many practical scenarios (Ahsan et al., 2024; Meng et al., 2020; Qadrini, 2022). One key concern is that SMOTE synthetically generates new data points by interpolating between existing minority class examples, which can introduce noise or artificial patterns that do not reflect the true underlying distribution

[4] The choice to employ these methods—such as class weighting and resampling—over alternative techniques for addressing class imbalance, including modified loss functions or generative adversarial networks (GANs), is grounded in their operational characteristics. Approaches like loss functions and GANs operate at the data level (see, e.g., Elreedy and Atiya, 2019), allowing them to be integrated with a wide range of classification algorithms without altering the model architecture or training procedure. However, such techniques are often more appropriate for domains with fundamentally different data structures and objectives, such as image classification tasks (Pan et al., 2024). In contrast, the selected methods are particularly well-suited to the tabular, structured data typical of energy poverty studies and offer a more interpretable and computationally efficient solution.

of the data. This can lead to overfitting, particularly when used in conjunction with high-capacity models. Additionally, in high-dimensional feature spaces, the synthetic samples may fall into regions dominated by the majority class, further degrading model performance. By contrast, undersampling techniques, which reduce the number of majority class examples, are often preferred for their simplicity and for maintaining the integrity of the original data. Undersampling can also help reduce computational cost and prevent models from being overwhelmed by majority class examples, making it a more robust and interpretable choice, especially when working with large datasets.

The imbalance in the dataset poses a significant challenge for machine learning models, especially in correctly identifying households in EP. Without addressing this imbalance, models may fail to serve the primary objective of predicting energy-poor households. By implementing techniques such as resampling, class weighting, and using appropriate evaluation metrics, we can improve the model's ability to handle the imbalance and accurately identify energy-poor households. In the following subsection, we report the evaluation metrics of each indicator, which are classified using the algorithms proposed and taking into consideration the imbalanced classes. We included all expenditure variables and household and dwelling characteristics from the HBS in the initial specifications, including the L1-regularisation for selecting relevant variables.

4.1 Random forest

Table 3 presents the classification metrics derived from applying the random forest algorithm to assess the three indicators of EP based on data from the Household Budget Survey (HBS). For each indicator, metrics are provided for i) the model estimated without adjustments for class imbalance and ii) the models estimated with the specified methods to address this issue.

The metrics reported in Table 3 reflect the effectiveness of a random forest classifier in predicting EP among Italian households, utilizing the three distinct indicators. Each indicator measures EP from different theoretical perspectives, and this table provides insight into the classifier's performance on each using accuracy, precision, recall, and F1-score as core metrics. The latter is a more appropriate metric for an imbalanced class dataset:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

The table also assesses four approaches to handling class imbalance: No Balancing, SMOTE, Class Weight Adjustment, and Undersampling.

ENPOV1 and ENPOV2 demonstrate high accuracy levels (0.96 and 0.92, respectively) without class balancing, suggesting that the model can effectively classify the dominant non-energy-poor class. However, a closer examination of recall (0.11 for ENPOV1 and 0.06 for ENPOV2) reveals that the classifier performs poorly in identifying households classified as energy-poor. This skew in performance, with a strong majority-class bias, confirms that accuracy alone may not be an adequate performance measure in the presence of class imbalance.

ENPOV3, on the other hand, maintains strong performance across all metrics (accuracy of 0.98, recall of 0.94, precision of 0.89, and F1-score of 0.91), even without class balancing. This suggests that ENPOV3 may provide a more discriminative measure for identifying EP in this dataset. It is plausible that the operationalization of EP in ENPOV3 captures characteristics more aligned with the dataset's features, facilitating the model's differentiation between energy-poor and non-energy-poor households.

Table 3 Classification metrics for random forest algorithms

Random forest																
	No balancing				Smote				Classe weight				Undersampling			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
ENPOV1	0.96	0.62	0.06	0.11	0.96	0.66	0.08	0.14	0.96	0.48	0.32	0.39	0.85	0.21	0.95	0.34
ENPOV2	0.92	0.57	0.11	0.19	0.92	0.63	0.10	0.18	0.92	0.61	0.08	0.15	0.79	0.27	0.89	0.41
ENPOV3	0.98	0.89	0.94	0.91	0.98	0.85	0.99	0.92	0.98	0.88	0.94	0.91	0.98	0.80	1.00	0.89

The class balancing methods (SMOTE, class weight adjustment, and undersampling) aim to mitigate the observed performance discrepancies, especially the low recall rates in ENPOV1 and ENPOV2.

By synthetically generating new samples of the minority class, SMOTE slightly improves recall for ENPOV1 and ENPOV2 (0.08 and 0.10, respectively) but shows limited success. Precision increases somewhat (from 0.57 to 0.63 for ENPOV2), but the overall improvement remains modest, reflecting a limited impact on true positive identification.

Assigning a higher weight to the minority class (energy-poor households) improves recall more substantially for ENPOV1 (from 0.06 to 0.32) while balancing precision. However, it slightly reduces the precision for ENPOV1 (from 0.66 to 0.48). Thus, this approach offers a more balanced performance for ENPOV1 but remains less impactful for ENPOV2.

Undersampling produces the most significant recall improvement across all indicators, especially for ENPOV2 (with recall rising to 0.89) and ENPOV1 (0.95). However, this method causes a decrease in overall accuracy for ENPOV2 (to 0.79) and ENPOV1 (to 0.85), an expected outcome due to the reduction of majority class instances. Despite the accuracy drop, the substantial recall increase indicates that undersampling can better address the class imbalance issue, particularly for applications where identifying energy-poor households is critical.

ENPOV1 and ENPOV2 struggle to achieve high recall values, suggesting potential limitations in capturing EP about the features used in the random forest model. While balancing techniques improve performance, the indicators still exhibit inconsistencies, with no approach to rectifying the recall deficit fully. This may suggest that these indicators lack sufficient differentiation for EP cases when viewed through household budget data.

ENPOV3 demonstrates robust, balanced performance with minimal need for class balancing. High recall and precision suggest that ENPOV3 may align more closely with the underlying structure of the data, possibly due to more targeted variable selection or a closer conceptual match to features indicative of EP. The consistency of ENPOV3 across all balancing methods further underscores its utility as a reliable predictor for EP.

4.2 XGBoosting

Table 4 presents the classification metrics of the XGBoost model.

ENPOV1 and ENPOV2 exhibit high accuracy scores (0.96 and 0.93, respectively) across balancing methods, but accuracy alone does not capture model performance effectively due to class imbalance. Both indicators display low recall without balancing adjustments (0.22 for ENPOV1 and 0.28 for ENPOV2), indicating that the model is less effective in identifying actual positive cases of energy-poor households, a critical factor for policy-focused applications.

ENPOV3, however, shows balanced performance across all metrics, maintaining high precision (0.85–0.86) and recall (0.92–0.96) without any class-balancing interventions. This suggests that ENPOV3 may be inherently better aligned with the data features, allowing the model to differentiate between energy-poor and non-energy-poor households with consistent minimal adjustment.

By oversampling the minority class, SMOTE marginally improves recall for ENPOV1 (0.25) and ENPOV2 (0.29) compared to the unbalanced model. However, precision remains similar, and the F1-score increases only slightly, indicating a limited impact. For ENPOV3, SMOTE's effect is minimal since the indicator already performs well without adjustment.

Table 4 Classification metrics for XGBoosting

XGBoosting																
	No balancing				Smote				Classe weight				Undersampling			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
ENPOV1	0.96	0.45	0.22	0.29	0.96	0.49	0.25	0.33	0.95	0.40	0.41	0.40	0.86	0.22	0.96	0.36
ENPOV2	0.93	0.54	0.28	0.37	0.93	0.53	0.29	0.38	0.91	0.43	0.44	0.44	0.79	0.25	0.87	0.39
ENPOV3	0.98	0.85	0.92	0.88	0.98	0.86	0.93	0.89	0.98	0.84	0.96	0.90	0.98	0.80	1	0.89

Assigning higher weights to the minority class (energy-poor households) results in a notable increase in recall for both ENPOV1 (0.41) and ENPOV2 (0.44). This method achieves a more balanced precision-recall trade-off, as reflected by improved F1-scores (0.44 for ENPOV2 and 0.40 for ENPOV1). This suggests that class weighting effectively reduces the recall deficit for these indicators while maintaining a reasonable precision level. ENPOV3's metrics remain high and stable, further indicating the robustness of this indicator.

The undersampling approach generates the highest recall values for ENPOV1 (0.96) and ENPOV2 (0.87) but reduces precision substantially (to 0.22 for ENPOV1 and 0.25 for ENPOV2). While this method maximizes the model's ability to detect energy-poor households, it results in many false positives, as the reduced precision shows. For ENPOV3, undersampling has a slight impact, lowering precision (0.80) while maximizing recall (1.0), which maintains the indicator's already high F1-score (0.89).

4.3 Stacked generalization

As was done for previous models, Table 5 reports the classification metrics for the stacked generalization.

ENPOV1 and ENPOV2 achieve high accuracy (0.96 and 0.93, respectively), similar to the results seen with other MLAs (see Tables 3 and 4). Similarly, both indicators have relatively low recall (0.23 for ENPOV2 and 0.17 for ENPOV1), indicating that while the model successfully predicts non-energy-poor households, it performs less effectively detecting actual positive cases of EP.

ENPOV3 again performs consistently well across metrics, with high accuracy (0.98), precision (0.87), and recall (0.92), even without balancing. This suggests that ENPOV3 inherently aligns with the feature structure of the data, enabling the model to distinguish energy-poor households without additional balancing effectively.

Each class balancing method has distinct effects on recall and precision, particularly for ENPOV1 and ENPOV2, which struggle more with class imbalance.

Applying SMOTE slightly improves recall for ENPOV1 (from 0.23 to 0.39) and ENPOV2 (from 0.17 to 0.25). The F1-score also improves marginally, reflecting a better balance between precision and recall for these indicators. However, the gains remain modest, suggesting that while SMOTE partially addresses the imbalance, it cannot capture the minority class to an ideal level.

Class Weight Adjustment appears beneficial for precision in ENPOV2 (increasing from 0.59 to 0.64) and retains high accuracy (0.93). However, recall improves only slightly (to 0.16 for ENPOV1 and 0.20 for ENPOV2). For ENPOV3, class weight adjustment has minimal effect as the indicator already performs strongly, reinforcing ENPOV3's robustness.

Undersampling yields the highest recall for ENPOV1 (0.95) and ENPOV2 (0.86) but at a notable cost to precision, especially in ENPOV1, where precision drops to 0.22. This drop suggests a trade-off inherent in undersampling: it increases recall by emphasizing minority cases, leading to a higher number of false positives (lower precision). The overall performance (F1-score) is best balanced here for ENPOV3, where undersampling improves recall to 1.0 without sacrificing overall accuracy or F1-score.

ENPOV1 and ENPOV2 demonstrate persistent limitations in recall across balancing techniques, particularly when compared to ENPOV3. While SMOTE and class weight adjustments lead to modest recall improvements, undersampling shows the most considerable impact, albeit with decreased precision. This pattern may indicate that ENPOV1 and ENPOV2, as operationalized, do not capture EP cases as distinctly within this dataset, leading

Table 5 Classification metrics for staked generalization algorithms

Staked generalization	No balancing				Smote				Classe weight				Undersampling			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
ENPOV1	0.96	0.58	0.17	0.27	0.96	0.44	0.25	0.32	0.96	0.59	0.16	0.25	0.86	0.22	0.95	0.36
ENPOV2	0.93	0.59	0.23	0.33	0.92	0.46	0.39	0.42	0.93	0.64	0.20	0.30	0.81	0.27	0.86	0.41
ENPOV3	0.98	0.87	0.92	0.89	0.98	0.83	0.97	0.90	0.98	0.86	0.96	0.90	0.98	0.80	1	0.89

to weaker performance. Further refinement or the inclusion of additional features aligned with the underlying construct of these indicators could potentially improve predictive accuracy for these cases.

ENPOV3, on the other hand, consistently achieves high accuracy, precision, and recall across all balancing methods, suggesting a solid alignment with the surveyed data. The classification metrics associated with ENPOV3 mean that it captures essential characteristics that reliably differentiate energy-poor households, making it the most effective indicator for this stacked model. Its robustness across all methods further indicates its potential as the preferred measure for policy applications where accurate identification of EP is critical.

4.4 Comparative analysis of machine learning algorithms across indicators

This subsection provides an integrated comparison of the 36 model configurations generated by combining three machine learning algorithms (Random Forest, XGBoost, and Stacked Generalization), three EP indicators (ENPOV1, ENPOV2, ENPOV3), and four class balancing strategies (No Balancing, SMOTE, Class Weighting, and Undersampling). The F1-score is used as the principal metric for evaluation, as it captures the trade-off between precision and recall, which are critical for accurately identifying energy-poor households in imbalanced datasets.

4.4.1 Summary of model performance

Figure 1 presents a heatmap of F1-scores for all model-indicator-balancing method combinations to facilitate interpretation. The rows represent the machine learning algorithms, while the columns correspond to each pairing of indicator and balancing strategy. Higher F1-scores (darker cells) indicate better model performance in identifying energy-poor households.

In addition to the heatmap, Table 6 reports the average F1-score for each algorithm-indicator pair across all balancing methods. This summary allows for a quick assessment of general trends and relative robustness of the indicators and algorithms.

Comparing the synthesis results included in Fig. 1 and Table 6, it appears that ENPOV3 is the most robust indicator across all models and balancing methods. ENPOV3 consistently yields the highest F1-scores, often exceeding 0.89, regardless of the classifier or class balancing strategy used. This indicates that the operationalization of EP in ENPOV3, possibly due to its inclusion of broader regional socioeconomic characteristics, aligns well with the

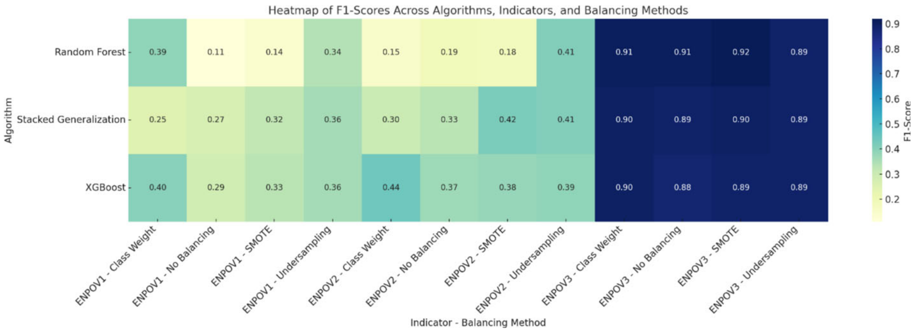

Fig. 1 Heatmap of F1-scores across all MLAs, indicator, and class balancing method combinations

Table 6 Average F1-scores across class balancing methods

Algorithm	ENPOV1	ENPOV2	ENPOV3
Random forest	0.24	0.23	0.91
Stacked Generalization	0.30	0.36	0.90
XGBoost	0.34	0.40	0.89

predictive features available in the HBS. Its performance remains high even in the absence of balancing, suggesting that ENPOV3 provides better separability between energy-poor and non-energy-poor households in the feature space.

However, ENPOV1 and ENPOV2 require balancing techniques to achieve acceptable recall and F1-scores. Both indicators perform poorly in unbalanced settings, with low F1-scores largely due to insufficient recall. However, the application of class weighting and, especially, undersampling, substantially improves recall, thus boosting F1-scores. Among the two, ENPOV2 exhibits slightly better average performance than ENPOV1. This suggests that while these indicators may capture valid dimensions of EP, their alignment with observed consumption and dwelling characteristics is weaker compared to ENPOV3.

Moreover, XGBoost and Stacked Generalization outperform Random Forest in handling imbalanced classification. Across ENPOV1 and ENPOV2, XGBoost achieves the highest average F1-scores (0.34 and 0.40, respectively), followed by Stacked Generalization (0.30 and 0.36). This pattern implies that more complex ensemble methods or boosting approaches are better suited for extracting signals from imbalanced and noisy household data. While Random Forest performs well with ENPOV3, its performance degrades significantly for the other two indicators, suggesting greater sensitivity to class imbalance and less adaptability across indicators.

These findings underscore the importance of indicator selection, algorithm choice, and proper handling of class imbalance in machine learning applications aimed at identifying EP. The evidence strongly favors ENPOV3 as the most effective indicator when used in combination with either XGBoost or Stacked Generalization. In contrast, ENPOV1 and ENPOV2 demand additional pre-processing steps (e.g., undersampling) and algorithmic tuning to reach comparable levels of performance. The comparative results also provide a comprehensive foundation for selecting optimal predictive strategies in future empirical research or policy deployment focused on EP detection.

5 Discussion

As is well known, EP is a complex and multidimensional issue that intersects with social, economic, and environmental factors, significantly impacting the well-being and quality of life of those affected. This study employs *machine learning algorithms* (MLAs) to predict EP in Italian households, offering a data-driven approach to identifying at-risk populations. The comparative analysis of three expenditure-based indicators – ENPOV1 (Low-Income, High-Costs), ENPOV2 (Modified LIHC), and ENPOV3 (*Regional Energy Poverty*) – provides critical insights into the effectiveness of MLAs for EP classification and methods for improving the performance with imbalanced class data.

The standout performance of ENPOV3 across all MLAs, particularly under conditions of class imbalance, demonstrates its robustness. This indicator effectively incorporates socioeconomic and regional disparities, aligning closely with the structure of data available from the HBS. Conversely, ENPOV1 and ENPOV2 exhibit inconsistent recall rates, indicating their limited utility in capturing the full spectrum of EP.

An in-depth analysis of the performance of MLAs reveals the following patterns:

- Random Forest algorithm performed well in precision but struggled with recall in ENPOV1 and ENPOV2, underscoring its bias towards the majority class without adequate balancing.
- After applying balancing techniques, XGBoost demonstrated superior recall improvement, particularly for ENPOV1 and ENPOV2. However, its high precision was achieved at the cost of reduced accuracy in undersampling scenarios.
- Stacked Models effectively leveraged the strengths of diverse base learners, achieving the highest F1 scores, especially for ENPOV3. This highlights the potential of ensemble methods in handling complex EP datasets.

The class imbalance represents one of the main critical challenges in predicting households in EP. The study shows that balancing techniques like SMOTE, class weighting, and undersampling significantly affect the performance of MLAs. Notably, while undersampling yielded high recall for ENPOV1 and ENPOV2, it compromised precision and accuracy. This trade-off underlines the importance of choosing the correct balance technique based on policy objectives – prioritizing comprehensive identification (recall) or minimizing false positives (precision).

ENPOV3's high performance across all metrics suggests that its formulation captures a broader and more realistic spectrum of EP. Its incorporation of regional and socioeconomic factors makes it a holistic tool for understanding EP, using readily available HBS data and offering to policymakers a reliable foundation for targeted interventions.

For the sake of simplicity, here are reported the Variable Importance Plots from Random Forest (Fig. 2) and XGBoost (Fig. 3).

The analysis of the variables' importance highlights that the feature "total household expenditures" stands out as the most important, with a significantly higher importance score than any other indicator. Other indicators, like "Tax credits for energy efficiency increase (regional average)", "Single-person household", "White collar household head", and "Size of dwelling (sqm)" follow with lower but notable importance scores. These features contribute to the model but are far less impactful than total household expenditures.

The remaining features have relatively small contributions, with diminishing importance scores as we move down the list (e.g., house provided by internet connection, etc.).

The dominance of total household expenditures in the feature importance plot is shared in the literature (e.g., Spandagos et al., 2023). Total household expenditures are a proxy for a family's income. Households with lower total expenditures may struggle to allocate sufficient funds for energy needs, increasing their likelihood of being in EP. As is well known, EP typically involves households that cannot afford adequate energy services. Since this feature aggregates all household spending, it likely captures the ability (or inability) to afford energy and other essential services.

Also, the tax incentives for improving residential energy efficiency contribute to capturing the nuanced financial dynamics of households, improving the ability of the MLAs to identify energy-poor families. The effectiveness of these policies has been studied in the literature, and it has been confirmed that they are effective in tackling EP (e.g., Faccioli et al., 2024; Martini, 2023).

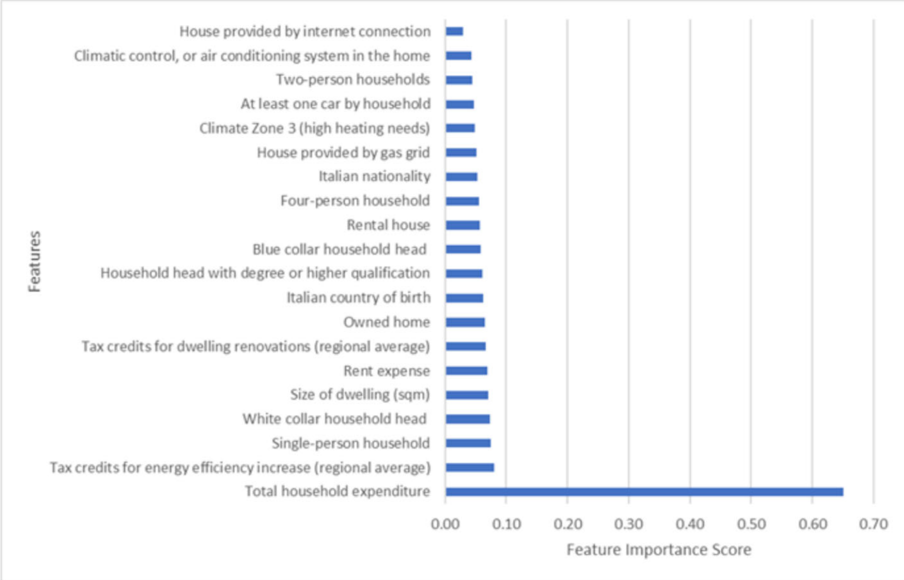

Fig. 2 Variance importance plot with random forest and smote

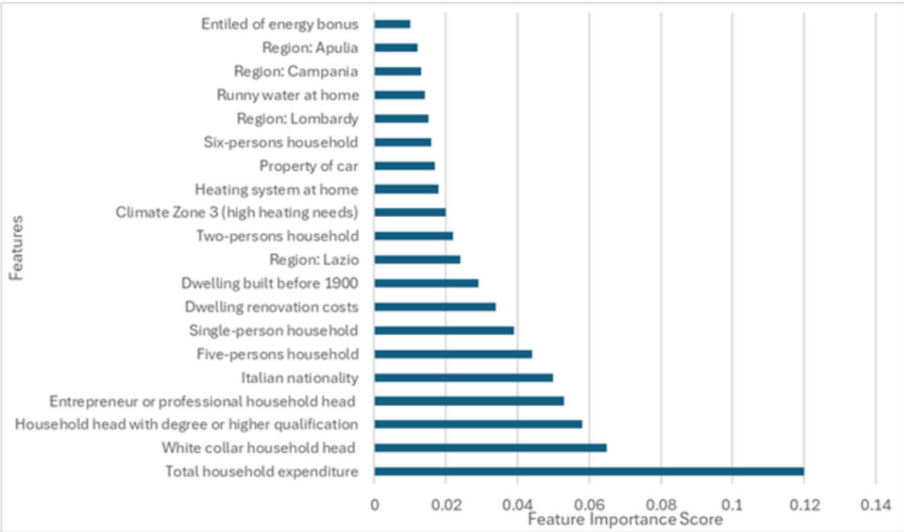

Fig. 3 Variance importance plot with XGBoost and undersampling

Other indicators, such as household size and dwelling size, may help indicate that households are under pressure due to limited income. The insights from this study highlight actionable pathways for policy improvement.

The robustness of ENPOV3 across various MLAs underscores its potential as a benchmark indicator for EP identification. It integrates region-specific data, such as local energy costs and housing characteristics, alongside socioeconomic indicators, allowing for the development of localized policies. High-risk areas, as identified by the model, should be prioritized for energy efficiency grants, retrofitting programs, and direct financial assistance. Such a comprehensive approach enables policymakers to identify households in EP more accurately than narrower indicators like ENPOV1 (focused solely on income-energy cost ratios) or ENPOV2 (which emphasizes affordability). Unlike other indicators, ENPOV3 consistently performs highly across all key evaluation metrics (accuracy, precision, recall, and F1-score). This reliability highlights its ability to capture the multifaceted nature of EP, including regional disparities, socioeconomic vulnerabilities, and variations in energy costs and housing conditions. Its predictive strength and alignment with real-world data make ENPOV3 a powerful tool for leading policy decisions and optimizing intervention strategies.

The high performance of ENPOV3 across the various MLAs demonstrates its robustness in various predictive environments. This reliability is crucial for policymakers who require dependable tools to inform resource allocation and intervention design. ENPOV3's performance holds steady even when addressing class imbalance, a common challenge in EP datasets, further reinforcing its practical applicability. Moreover, ENPOV3's superior recall ensures that fewer energy-poor households are overlooked, addressing a critical gap in policy targeting. By accurately identifying a broader spectrum of energy-poor households, policymakers can better design and deploy targeted measures such as energy subsidies, retrofitting programs, and financial aid.

Policymakers should integrate ENPOV3 into national and regional policy frameworks to facilitate dynamic, data-driven resource allocation in the fight against EP. At the national level, ENPOV3 can inform the design and evaluation of overarching policy instruments, such as energy subsidy programs, renovation incentives, and energy efficiency regulations. At the regional level, it enables the development of context-specific interventions, such as increased support for heating in colder climates or targeted retrofitting initiatives in areas with aging housing stock, based on localized needs and vulnerabilities.

Policymakers should integrate ENPOV3 into national and regional policy workflows to enable dynamic, data-driven resource allocation to tackle EP. At the national level, ENPOV3 can guide overarching policy frameworks, including energy subsidy schemes, renovation bonuses, and energy efficiency mandates. Regionally, ENPOV3 can help local governments design interventions that reflect specific regional needs, such as higher heating requirements in colder areas or targeted retrofitting in regions with older housing stock.

In an increasingly interconnected and rapidly changing world, effectively addressing complex societal challenges like EP necessitates the adoption of advanced analytical frameworks. Traditional policy approaches often fail to capture the systemic, multidimensional nature of such issues. To address these limitations, recent research advocates for integrating complexity science and data science into a unified analytical paradigm known as complexity data science (Emmert-Streib et al., 2024). This interdisciplinary framework combines the theoretical insights of complex systems with the empirical strengths of data-driven inference to provide a more comprehensive understanding of socio-technical phenomena (Lu & Ren, 2023).

Within this framework, ENPOV3 can serve as a foundational tool, enabling real-time refinement of targeting strategies through continuous model updates based on evolving data inputs, such as fluctuations in energy prices or changes in household income. This adaptive capability ensures that scarce resources, including grants and subsidies, are allocated efficiently and equitably, maximizing their impact on those most in need.

6 Conclusion and policy implications

This study highlights the transformative potential of *machine learning algorithms* (MLAs) in addressing energy poverty (EP) by providing policymakers with robust, data-driven tools to identify and support vulnerable households. By leveraging the three expenditure-based EP indicators proposed by Hills (2011) (ENPOV1), Faiella and Lavecchia (2014) (ENPOV2), and Betto et al. (2020) (ENPOV3), alongside detailed data from the *Household Budget Survey* (HBS), this research offers a comprehensive framework for understanding and mitigating EP in Italy. It contributes to the growing research stream on EP by applying advanced MLAs to the Italian context, an issue that has received limited attention in data-driven EP research. By leveraging ensemble models, we achieve high predictive accuracy and uncover key predictors that inform targeted policy interventions. The novelty of this work lies in its integration of predictive analytics with interpretability, enabling a deeper understanding of the multidimensional drivers of EP.

ENPOV3 is the most effective among the indicators analyzed, consistently delivering high performance across all machine learning models and evaluation metrics. Its ability to incorporate regional and socioeconomic disparities provides a holistic view of EP, aligning well with real-world complexities. In contrast, while applicable in specific contexts, the other two indicators demonstrated limitations in recall and sensitivity to class imbalance, indicating that they may miss specific subsets of energy-poor households.

The study's comparative analysis of the MLAs, Random Forest, XGBoost, and Stacked Generalization further underscores the value of ensemble approaches. Stacked models, in particular, proved adept at synthesizing diverse learners, achieving the highest F1-scores and balancing the trade-offs between precision and recall. This outcome reinforces the importance of leveraging advanced ML techniques to enhance predictive accuracy in complex, multidimensional policy areas like EP.

The superior performance of ENPOV3 makes it a powerful tool for guiding targeted interventions. Its adoption as a primary indicator can significantly enhance the precision of EP mitigation strategies, ensuring that resources are allocated effectively to regions and households that suffer most from EP.

Based on readily available data, the results suggest to policymakers to integrate MLAs into policy frameworks, moving beyond static, one-size-fits-all solutions. Dynamic, real-time information allows for the continuous refinement of predictive models, allowing policymakers to adapt their strategies in response to emerging challenges such as economic shocks or energy price fluctuations. This adaptability ensures timely and effective responses to mitigate EP and its evolving drivers.

6.1 Policy implications and future research

While fiscal incentives for energy-efficient renovations have proven beneficial, their impact can be uneven across household profiles. Complementary measures, such as subsidies for energy-efficient appliances and direct financial support for low-income households, should be integrated to address residual energy burdens. These measures can ensure comprehensive support for families transitioning out of EP. Moreover, integrating indicators into national and regional policies enables the design of interventions tailored to the specific needs of different regions. By capturing local disparities in energy access and socioeconomic conditions, ENPOV3 helps prioritize areas with higher EP prevalence. Regional authorities can use this data to implement localized programs, such as energy retrofitting or targeted subsidies, ensuring efficient allocation of resources.

Collaborative initiatives could combine financial assistance with infrastructure improvements and educational campaigns to promote energy efficiency and conservation. This holistic approach could maximize the impact of interventions and foster a sustainable energy transition.

Future data collection efforts must address current limitations to maximize the potential of MLAs. Expanding the HBS to include more granular data on energy consumption, housing quality, and regional climatic variations will improve model accuracy and robustness. This can also be useful to reduce over-reliance on a single feature like the total household expenditures (spese_comp). Additionally, oversampling vulnerable households in surveys will help mitigate class imbalance, a persistent challenge in predictive modeling.

EP is a growing challenge with far-reaching implications for social equity, public health, and sustainable development. This study demonstrates that MLAs applied to known and established statistical survey data, implemented by the Istat, offer a powerful means to address this issue, enabling more accurate identification of at-risk households and more effective policy responses to alleviate EP and support a just and sustainable energy transition.

Although the results provide clear indications of the applicability of MLAs on energy-poor household classification, it should be considered that they are based on basic model applications. The results should be compared with those from hybrid models. Furthermore, we have to include spatial and longitudinal analyses to identify patterns and trends. Moreover, the models should be replicated based on similar surveys done in other countries by looking for a common framework in the EU-SILC survey.

Beyond their predictive performance, the MLAs employed in this study offer broader implications for both research and policy-makers in the EU. By comparing multiple algorithms and indicators, we demonstrate that MLAs can support public authorities in developing more accurate tools for identifying energy-poor households, an essential step for designing effective, targeted support schemes under the EU's Energy Performance of Buildings Directive and "Fit for 55" framework.

The use of these techniques also raises important ethical and legal considerations. As integrating administrative and survey-based datasets becomes more feasible, particularly through national statistics and open government initiatives, ensuring compliance with data protection laws (e.g., General Data Protection Regulation) and maintaining transparency and fairness in algorithmic decision-making becomes paramount. Risks related to biased models, misuse of sensitive attributes, or opaque classification processes must be managed carefully, especially when decisions may affect eligibility for support measures.

This approach could be extended to other areas of social and economic vulnerability, including housing instability, healthcare access, or food insecurity. These domains also suffer from data imbalance and multidimensional causality, making them suitable candidates for ML-based classification frameworks informed by national microdata.

Future research should explore integrating other data sources, such as utility billing data, geospatial information, and real-time energy use patterns. Moreover, it could extend this framework to incorporate temporal dynamics, explore causal inference using counterfactual ML methods, or apply the model to other European countries to support comparative policy analysis.

Integrating these data streams can further improve the predictive power of MLAs and provide insights into the structural factors of EP. Although this study focuses on Italy, the adopted methodology can serve as a model for wider applications across the EU. A harmonised, pan-European framework for measuring and mitigating EP, based on MLAs, could foster greater policy coherence and improve outcomes for vulnerable households across EU countries.

Data availability Data are public from the ISTAT database.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbas, K., Butt, K., Xu, D., Ali, M., Baz, K., Kharl, S. H., & Ahmed, M. (2022). Measurements and determinants of extreme multidimensional energy poverty using machine learning. *Energy, 251*, 123977. https://doi.org/10.1016/j.energy.2022.123977

Ahsan, M. M., Ali, M. S., & Siddique, Z. (2024). Enhancing and improving the performance of imbalanced class data using novel GBO and SSG: A comparative analysis. *Neural Networks, 173*, 106157. https://doi.org/10.1016/j.neunet.2024.106157

Al Kez, D., Foley, A., Abdul, Z. K., & Del Rio, D. F. (2024a). Energy poverty prediction in the United Kingdom: A machine learning approach. *Energy Policy, 184*, 113909. https://doi.org/10.1016/j.enpol.2023.113909

Al Kez, D., Foley, A., Lowans, C., & Del Rio, D. F. (2024b). Energy poverty assessment: Indicators and implications for developing and developed countries. *Energy Conversion and Management, 307*, 118324. https://doi.org/10.1016/j.enconman.2024.118324

Babayomi, O. O., Dahoro, D. A., & Zhang, Z. (2022) Affordable clean energy transition in developing countries: Pathways and technologies. *Iscience*, *25*(5).

Banerjee, R., Mishra, V., & Maruta, A. A. (2021). Energy poverty, health and education outcomes: Evidence from the developing world. *Energy Economics, 101*, 105447.

Bardazzi, R., Gastaldi, F., Iafrate, F., Pansini, R. V., Pazienza, M. G., & Pollastri, C. (2024). Inflation and distributional impacts: Have mitigation policies been successful for vulnerable and energy poor households? *Energy Policy, 188*, 114082.

Barnard, L. T., Howden-Chapman, P., Clarke, M., Ludolph, R. (2018) Report of the systematic review on the effect of indoor cold on health. Geneva, Switzerland

Basri, A., & Arif, M. (2021). Classification of seizure types using random forest classifier. *Advances in Science and Technology Research Journal, 15*(3), 167–178. https://doi.org/10.12913/22998624/140542

Betto, F., Garengo, P., & Lorenzoni, A. (2020). A new measure of Italian hidden energy poverty. *Energy Policy, 138*, 111237.

Billio, M., Casarin, R., Costola, M., & Veggente, V. (2024). Learning from experts: Energy efficiency in residential building. *Energy Economics, 136*, 107650. https://doi.org/10.1016/j.eneco.2024.107650

Boardman B. (1991). Fuel poverty: from cold homes to affordable Warmth. Belhaven Press.

Bouzarovski, S., Thomson, H., Cornelis, M. (2021). Confronting Energy Poverty in Europe: A Research and Policy Agenda. *Energies, 14*, 858. https://doi.org/10.3390/en14040858

Bramley, G., Fitzpatrick, S., Liddell, C., Webb, J. (2017). A new definition of fuel poverty in scotland: a review of recent evidence. Edinburgh, United Kingdom.

Breiman, I. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brownlee, J. (2020). How to develop a light gradient boosted machine (LightGBM) ensemble. https://machinelearningmastery.com/light-gradient-boosted-machine-lightgbm-ensemble/. Accessed September 28, 2024.

Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science, 22*(4), 477–505. https://doi.org/10.1214/07-STS242
Carfora, A., & Scandurra, G. (2024). Boosting green energy transition to tackle energy poverty in Europe. *Energy Research & Social Science, 110*, 103451.
Casarin, R., Facchinetti, A., Sorice, D., & Tonellato, S. (2021). In M. Z. Abedin, M. K. Hassam, P. Hajek, & M. M. Uddin (Eds.), *Decision trees and random forests.* Routledge, Taylor & Francis.
Castaño-Rosa, R., Sherriff, G., Solís-Guzmán, J., & Marrero, M. (2020). The validity of the index of vulnerable homes: Evidence from consumers vulnerable to energy poverty in the UK. *Energy Sources, Part B: Economics, Planning, and Policy, 15*(2), 72–91.
Chang, V., Xu, Q. A., Akinloye, S. H., Benson, V., & Hall, K. (2024). Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study. *Annals of Operations Research*. https://doi.org/10.1007/s10479-024-06134-x
Chen, T., Guestrin, C., (2016). HGBoost: A scalable tree boosting system. In: Proocedings of the 22nd Acm Sigkdd international conferene on knowledge discovery and data mining, pp. 785–794.
Chengsheng, T. U., Huacheng, L. I. U., & Bing, X. U. (2017). AdaBoost typical algorithm and its application research. *MATEC Web of Conferences, 139*, 00222. https://doi.org/10.1051/matecconf/201713900222
Chowdhury, M. S. (2024). Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. *Environmental Challenges, 14*, 100800. https://doi.org/10.1016/j.envc.2023.100800
Dobbins, A., Nerini, F. F., Deane, P., & Pye, S. (2019). Strengthening the EU response to energy poverty. *Nature Energy*. https://doi.org/10.1038/S41560-018-0316-8
Dogan, E., Madaleno, M., & Taskin, D. (2021). Which households are more energy vulnerable? Energy poverty and financial inclusion in Turkey. *Energy Economics*. https://doi.org/10.1016/j.eneco.2021.105306
Dubois, U. (2012). From targeting to implementation: The role of identification. *Energy Policy, 49*, 107–115. https://doi.org/10.1016/j.enpol.2011.11.087
Elreedy, D., & Atiya, A. F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences, 505*, 32–64. https://doi.org/10.1016/j.ins.2019.07.070
Emmert-Streib, F., Cherifi, H., Kaski, K., Kauffman, S., & Yli-Harja, O. (2024). Complexity data science: A spin-off from digital twins. *PNAS Nexus*. https://doi.org/10.1093/pnasnexus/pgae456
Eurostat (2023). Population unable to keep home adequately warm by poverty status https://ec.europa.eu/eurostat/databrowser/view/SDG_07_60/default/table?lang=en
Faccioli, M., Lucarelli, C., Mazzoli, C., et al. (2024). What are the policy drivers to undertake green retrofitting investments? The role of tax incentives and communication in Italian households' decision-making. *Italian Economic Journal*. https://doi.org/10.1007/s40797-024-00283-5
Faiella, I., & Lavecchia, L. (2015). Energy poverty in Italy. *Polit. Econ. J. Econ. Policy, XXXI*, 27–76.
Faiella, I., & Lavecchia, L. (2021). Energy poverty. How can you fight it, if you can't measure it? *Energy and Buildings, 233*, 110692.
Gatera, A., Kuradusenge, M., Bajpai, G., Mikeka, C., & Shrivastava, S. (2023). Comparison of random forest and support vector machine regression models for forecasting road accidents. *Scientific African, 21*, E01739. https://doi.org/10.1016/j.sciaf.2023.e01739
Hasheminasab, H., Streimikiene, D., & Pishahang, M. (2023). A novel energy poverty evaluation: Study of the European Union countries. *Energy, 264*, 126157.
Henry, C. L., Baker, J. S., Shaw, B. K., Kondash, A. J., Leiva, B., Castellanos, E., Wade, C. M., Lord, B., Van Houtven, G., & Redmon, J. H. (2021). How will renewable energy development goals affect energy poverty in Guatemala? *Energy Economics, 104*, 105665.
Herrero, S. T. (2017). Energy poverty indicators: A critical review of methods. *Indoor and Built Environment, 26*(7), 1018–1031. https://doi.org/10.1177/1420326X17718054
Hills, J. (2011). Fuel poverty: the problem and its measurement, CASE report, 69. Department for Energy and Climate Change, London, UK.
Hussain, S. A., Razi, F., Hewage, K., & Sadiq, R. (2023). The perspective of energy poverty and 1st energy crisis of green transition. *Energy, 275*, 127487. https://doi.org/10.1016/j.energy.2023.127487
IEA (2024) Geopolitical tensions are laying bare fragilities in the global energy system, reinforcing need for faster expansion of clean energy from WEO - IEA - 16 ott 2024 https://www.iea.org/news/geopolitical-tensions-are-laying-bare-fragilities-in-the-global-energy-system-reinforcing-need-for-faster-expansion-of-clean-energy.
Jean, N., Burke, M., Xie, M., Alampay Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894.

Kahouli, S., & Okushima, S. (2021). Regional energy poverty reevaluated: A direct measurement approach applied to France and Japan. *Energy Economics*. https://doi.org/10.1016/j.eneco.2021.105491

Kashour, M., & Jaber, M. M. (2024). Revisiting energy poverty measurement for the European Union. *Energy Research & Social Science, 109*, 103420. https://doi.org/10.1016/j.erss.2024.103420

Katoch, O. R., Sharma, R., Parihar, S., & Nawaz, A. (2024). Energy poverty and its impacts on health and education: A systematic review. *International Journal of Energy Sector Management, 18*(2), 411–431.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*, 689–702.

Kyprianou, I., Serghides, D. K., Varo, A., Gouveia, J. P., Kopeva, D., Murauskaite, L. (2019). Energy poverty policies and measures in 5 EU countries: A comparative study. *Energy and Buildings, 196*, 46–60. https://doi.org/10.1016/j.enbuild.2019.05.003

Leimbach, M., & Giannousakis, A. (2019). Burden sharing of climate change mitigation: Global and regional challenges under shared socioeconomic pathways. *Climatic Change, 155*(2), 273–291.

Ljung, L., Andersson, C., Tiels, K., & Schön, T. B. (2020). Deep learning and system identification. *IFAC-PapersOnLine, 53*(2), 1175–1181.

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

López-García, A., Blasco-Blasco, O., Liern-García, M., & Parada-Rico, S. E. (2023). Early detection of students' failure using Machine learning techniques. *Operations Research Perspectives, 11*, 100292. https://doi.org/10.1016/j.orp.2023.100292

López-Vargas, a, Ledezma-Espino, a, & Sanchis-de-Miguel, a. (2022). Methods, data sources and applications of the artificial intelligence in the energy poverty context: A review. *Energy and Buildings, 268*, 112233. https://doi.org/10.1016/j.enbuild.2022.112233

Lu, S., & Ren, J. (2023). A comprehensive review on energy poverty: Definition, measurement, socioeconomic impact and its alleviation for carbon neutrality. *Environment, Development and Sustainability*. https://doi.org/10.1007/s10668-023-04143-7

Lu, S., & Ren, J. (2024). A global spatial-temporal energy poverty assessment and social impacts analysis. *International Journal of Energy Research, 2024*(1), 8247272.

Martini, C. (2023). Assessing ecobonus as energy poverty mitigation policy: Is energy efficiency for all? In R. Bardazzi & M. G. Pazienza (Eds.), *Vulnerable households in the energy transition: Studies in energy, resource and environmental economics.* Cham: Springer. https://doi.org/10.1007/978-3-031-35684-1_9

Meng, C., Zhou, L., & Liu, B. (2020). A case study in credit fraud detection with SMOTE and XGboost. *Journal of Physics: Conference Series, 1601*(5), 052016.

Middlemiss, L. (2022). Who is vulnerable to energy poverty in the Global North, and what is their experience? *Wiley Interdisciplinary Reviews: Energy and Environment, 11*(6), Article e455.

Mo, H., Sun, H., Liu, J., & Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. *Energy and Buildings, 205*, 109564.

Mukelabai, M. D., Wijayantha, K. G. U., & Blanchard, R. E. (2023). Using machine learning to expound energy poverty in the global south: Understanding and predicting access to cooking with clean energy. *Energy and AI, 14*, 100290. https://doi.org/10.1016/j.egyai.2023.100290.

Nussbaumer, P., Bazilian, M., & Modi, V. (2012). Measuring energy poverty: Focusing on what matters. *Renewable & Sustainable Energy Reviews, 16*(1), 231–243. https://doi.org/10.1016/j.rser.2011.07.150

Nussbaumer, P., Nerini, F., Onyeji, I., & Howells, M. (2013). Global insights based on the multidimensional energy poverty index (MEPI). *Sustainability, 5*(5), 2060–2076. https://doi.org/10.3390/su5052060

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S., Garibay, I., Grieman, K., et al. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human–Computer Interaction, 39*(3), 391–437. https://doi.org/10.1080/10447318.2022.2153320

Palma, P., Gouveia, J. P. (2022). Bringing energy poverty research into local practice: Exploring subnational scale analyses. Lisbon, Portugal.

Pan, T., Pedrycz, W., Yang, J., & Wang, J. (2024). An improved generative adversarial network to oversample imbalanced datasets. *Engineering Applications of Artificial Intelligence, 132*, 107934. https://doi.org/10.1016/j.engappai.2024.107934

Panão, M. J. N. O. (2021). Lessons learnt from using energy poverty expenditure-based indicators. *Energy & Buildings, 242*, 110936. https://doi.org/10.1016/j.enbuild.2021.110936

Papada, L., & Kaliampakos, D. (2022). Exploring energy poverty indicators through artificial neural networks. In M. Pandit, M. K. Gaur, P. S. Rana, & A. Tiwari (Eds.), *Artificial intelligence and sustainable computing algorithms for intelligent systems.* Singapore: Springer. https://doi.org/10.1007/978-981-19-1653-3_18

Primc, K., & Erker, R. (2020). Social policy or energy policy? Time to reconsider energy poverty policies. *Energy for Sustainable Development*. https://doi.org/10.1016/j.esd.2020.01.001

Puttanapong, N., Martinez Jr, A., Bulan, J. A. N., Addawe, M., Durante, R. L., & Martillan, M. (2022). Predicting poverty using geospatial data in Thailand. *ISPRS International Journal of Geo-Information, 11*(5), 293. https://doi.org/10.3390/ijgi11050293

Pye, S., Dobbins, A., Baffert, C., Brajkovi´c, J., Grgurev, I., De Miglio, R., Deane, P. (2015). Energy poverty and vulnerable consumers in the energy sector across the EU: Analysis of Policies and Measures.

Qadrini, L. (2022). Handling unbalanced data with SMOTE AdaBoost. *Jurnal Mantik, 6*(2), 2332–2336.

Rademaekers, K. et al., (2016). Selecting indicators to measure energy poverty. Rotterdam, Netherlands. https://ec.europa.eu/energy/sites/ener/files/documents/ Selecting Indicators to Measure Energy Poverty.pdf.

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble classification and regression-recent developments, applications and future directions. *IEEE Comput. Intell. Mag., 11*(1), 41–53.

Santangelo, A., Yan, D., Feng, X., & Tondelli, S. (2019). Renovation strategies for the Italian public housing stock: Applying building energy simulation and occupant behaviour modelling to support decision-making process. *Energy and Buildings, 167*, 269–280. https://doi.org/10.1016/j.enbuild.2018.02.028

Sareen, S., Thomson, H., Tirado Herrero, S., Gouveia, J. P., Lippert, I., & Lis, A. (2020). European energy poverty metrics: Scales, prospects and limits. *Global Transitions, 2*, 26–36. https://doi.org/10.1016/j.glt.2020.01.003

Scandurra, G., Carfora, A., & Thomas, A. (2024). Exploring the effect of energy transition on energy poverty in the European Union: Empirical evidence. *Energy Sources Part B: Economics, Planning, and Policy*. https://doi.org/10.1080/15567249.2024.2419437

Spandagos, C., Reanos, M. A. T., & Lynch, M.-A. (2023). Energy poverty prediction and effective targeting for just transitions with machine learning. *Energy Economics, 18*, 107131. https://doi.org/10.1016/j.eneco.2023.107131

Stojilovska, A., Guyet, R., Mahoney, K., Gouveia, J.P., Castaño-Rosa, R., Živčič, L., Barbosa, R. and Tkalec, T. (2022). Energy poverty and emerging debates: Beyond the traditional triangle of energy poverty drivers, Energy Policy, Elsevier, vol. 169(C).

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences, 513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Thomson, H., & Snell, C. (2013). Quantifying the prevalence of fuel poverty across the European Union. *Energy Policy, 52*, 563–572.

Thomson, H., Bouzarovski, S., & Snell, C. (2017). Rethinking the measurement of energy poverty in Europe: A critical analysis of indicators and data. *Indoor and Built Environment, 26*(7), 879–901. https://doi.org/10.1177/1420326X17699260

Thomson, H., Bouzarovski, S. (2019). Addressing energy poverty in the European Union: State of play and action, EU energy poverty observatory, Manchester

Tovar, M. A. (2021). Fuel for poverty: A model for the relationship between income and fuel poverty. Evidence from Irish microdata. *Energy Policy*. https://doi.org/10.1016/j.enpol.2021.112444

Valeria, A. (2024). Energy poverty in EU: Using regional climatic conditions and incidence of electricity prices to map vulnerability areas across 214 NUTS2 European regions. *World Development Sustainability, 4*, 100146.

Wang, B., Fan, S.-d, Jiang, P., Zhu, H.-h, Xiong, T., Wei, W., & Fang, Z.-l. (2020). A novel method with stacking learning of data-driven soft sensors for mud concentration in a cutter suction dredger. *Sensors, 20*, 6075. https://doi.org/10.3390/s20216075

Widuto, A. (2023). Energy poverty in the EU, European parliamentary research service (PE 733.583), https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733583/EPRS_BRI(2022)733583_EN.pdf

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259.

Wu, Z., Zhou, C., Xu, F., & Lou, W. (2020). A CS-AdaBoost-BP model for product quality inspection. *Annals of Operations Research*. https://doi.org/10.1007/s10479-020-03798-z

Zhang, Y., Hyder, M., Baloch, Z. A., Qian, C., & Saydaliev, H. B. (2022). Nexus between oil price volatility and inflation: Mediating nexus from exchange rate. *Resources Policy, 79*, 102977.

Zhao, J., Dong, K., Dong, X., & Shahbaz, M. (2022). How renewable energy alleviate energy poverty? A global analysis. *Renewable Energy, 186*, 299–311.