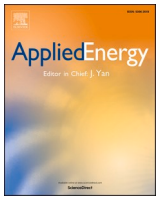
Applied Energy 366 (2024) 123314

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning

Hamidreza Eskandari [a,*], Hassan Saadatmand [b], Muhammad Ramzan [c,d], Mobina Mousapour [e]

[a] School of Management, Swansea University, Swansea, United Kingdom
[b] Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Faculty of Management and Administrative Sciences, University of Sialkot, Punjab, Pakistan
[d] Adnan Kassar School of Business, Lebanese American University, Beirut, Lebanon
[e] School of Industrial and Systems Engineering, College of Engineering, University of Tehran, Tehran, Iran

HIGHLIGHTS

- Innovative framework integrates FS and ML for accurate energy consumption forecasting.
- Strategic fusion: 3 FS approaches, 5 interpretable ML models with Shapley Explanations.
- Meticulous FS: meteorological, socioeconomic, historical features enhance model robustness.
- Novel ensemble filter synthesizes outcomes, ensuring interpretable forecasting results.

ARTICLE INFO

Keywords:
Energy consumption forecasting
Interpretable machine learning
Ensemble feature selection
Wrapper feature selection
Shapley analysis

ABSTRACT

The study presents a novel framework integrating feature selection (FS) and machine learning (ML) techniques to forecast inland national energy consumption (EC) in the United Kingdom across all energy sources. This innovative framework strategically combines three FS approaches with five interpretable ML models using Shapley Additive Explanations (SHAP), with the dual goal of enhancing accuracy and transparency in EC predictions. By meticulously selecting the most pertinent features from diverse features—including meteorological conditions, socioeconomic parameters, and historical consumption patterns of different primary fuels—the proposed framework enhances the robustness of the forecasting model. This is achieved through benchmarking three FS approaches: ensemble filter, wrapper, and a hybrid ensemble filter-wrapper. In addition, we introduce a novel ensemble filter FS, synthesizing outcomes from multiple base FS methods to make well-informed decisions about feature retention. Experimental results underscore the efficacy of integrating both wrapper and ensemble filter-wrapper FS approaches with interpretable ML models, ensuring the forecasting process remains comprehensible and interpretable while utilizing a manageable number of features (four to eight). In addition, experimental results indicate that different feature subsets are usually selected for each combined FS approach and ML model. This study not only demonstrates the framework's capability to provide accurate forecasts but also establishes it as a valuable tool for policymakers and energy analysts.

1. Introduction

1.1. Background

The worldwide community has united to pass an intricate framework of rules and policies meant to mitigate the damaging effects of GHGs under the overarching umbrella of the United Nations (UN). For a multitude of researchers, the phenomenon of global warming finds an inextricable link with the climatic expansion of the global economy and the accompanying increase in GHG emissions due to increase in energy consumption (EC) [1].

The energy landscape of the UK has been evolving in profound ways

* Corresponding author.
E-mail address: Hamidreza.Eskandari@Swansea.ac.uk (H. Eskandari).

https://doi.org/10.1016/j.apenergy.2024.123314
Received 23 December 2023; Received in revised form 12 April 2024; Accepted 22 April 2024
Available online 29 April 2024

for decades. Although historically reliant on fossil fuels, there has been a significant transition in recent years toward cleaner, more sustainable energy sources, as depicted in Fig. 1. As per the UK's National Grid, 2020 was a record-breaking year for renewable energy generation, with zero-carbon electricity constantly reaching 50% in numerous months of 2022. Coal now accounts for only 1.5% of electrical generation, down from 43% in 2012. Wind power records were broken, and carbon intensity reached historic lows in February 2022. Zero-carbon sources maintained their lead over conventional fossil fuel generation in 2022, contributing 48.5% of the total electricity consumption, while gas and coal power stations accounted for 40%. The UK's commitment to a low-carbon future remains unwavering. However, underlying this apparently smooth transmit to more sustainable energy sources is also source of a significant concern- the affordability of energy. Energy prices have risen in recent decades because of growing technology and environmental concern, resulting in what has been known as "energy poverty". According to UK's Department of Energy Security and Net Zero, in 2022, approximately 13.4% of households in England, totaling 3.26 million, were classified as experiencing fuel poverty under the Low Income Low Energy Efficiency (LILEE) metric. This marked an increase from 13.1% in 2021, affecting 3.16 million households. This problem creates concerns for individuals and families alike about their capacity to effectively heat their houses during hard winters, the availability of energy-intensive amenities, and the financial burden that energy bills impose on household budgets.

At the same time, it emphasizes the importance for governments, utilities, and energy suppliers to find a balance between sustainable energy practices and maintaining energy availability for everyone. This challenge also introduces an asymmetry in the pattern of energy demand in the UK, leading to disruptions in policy planning. To address this issue, precise forecast of consumption trends using robust approaches is required. Hence, we leverage a range of novel Machine Learning (ML) models, including multiple linear regression (MLR) [2], SVM-linear [3,4], SVM-RBF [5], Gaussian Process regression [4], and LSBoost [6], to enhance the accuracy and robustness of our EC forecasts for each sector in the UK. For that reason, the prediction of total EC is not merely an academic exercise; it serves as a critical tool for guiding decision-makers, optimizing resource allocation, and securing the energy future of the country.

1.2. Novelty and contributions

Our study marks a multifaceted contribution to the domain of ML based EC forecasting including:

- This study is a novel in advancing the use of monthly EC data at the national level, an intricate process given the typically restricted accessibility of such data for all energy sources, notably beyond the scope of electric EC. In contrast, other studies (e.g., [7,8,3]) exception of Li et al. [9], have mainly relied on annual national-level EC data characterized by a restricted number of observations, which potentially raises questions about the reliability of their findings. In pursuit of advancing the robustness and precision of our study, we leverage the potency of high-frequency monthly data.
- This approach not only enhances the accuracy of our ML-based forecasting but also contributes fresh insights to its features. Our research delves into multi-source EC statistics in the UK, spanning across all sectors at the national level. What sets this study apart is its comprehensive consideration of various features, encompassing energy sources, economic and demographic factors, climate variables, and an array of resource and production factors. This stands in stark contrast to prior research, which often relied on a limited set of pre-selected features without a robust justification.
- A novel generalizable ensemble FS approach is introduced, in which we iteratively eliminate the least important features in each filter FS method and employ a combined scoring and voting scheme for final ranking and selection of features. This ensemble FS process carefully uncovers and eliminates insignificant and redundant features, resulting in a refined and sparse subset that enhances the interpretability of complex ML models.
- Our study distinguishes itself through a meticulous and comparative FS analysis. This innovative framework integrates three distinct FS approaches —ensemble filter, wrapper, and ensemble filter-wrapper— with five ML models, providing a comprehensive comparative analysis.
- Our research makes a noteworthy contribution by proposing a robust and interpretable ML-based forecasting structure, depicted in Fig. 2. Notably, our framework strives for two critical objectives: improved prediction accuracy and enhanced interpretability of black-box ML

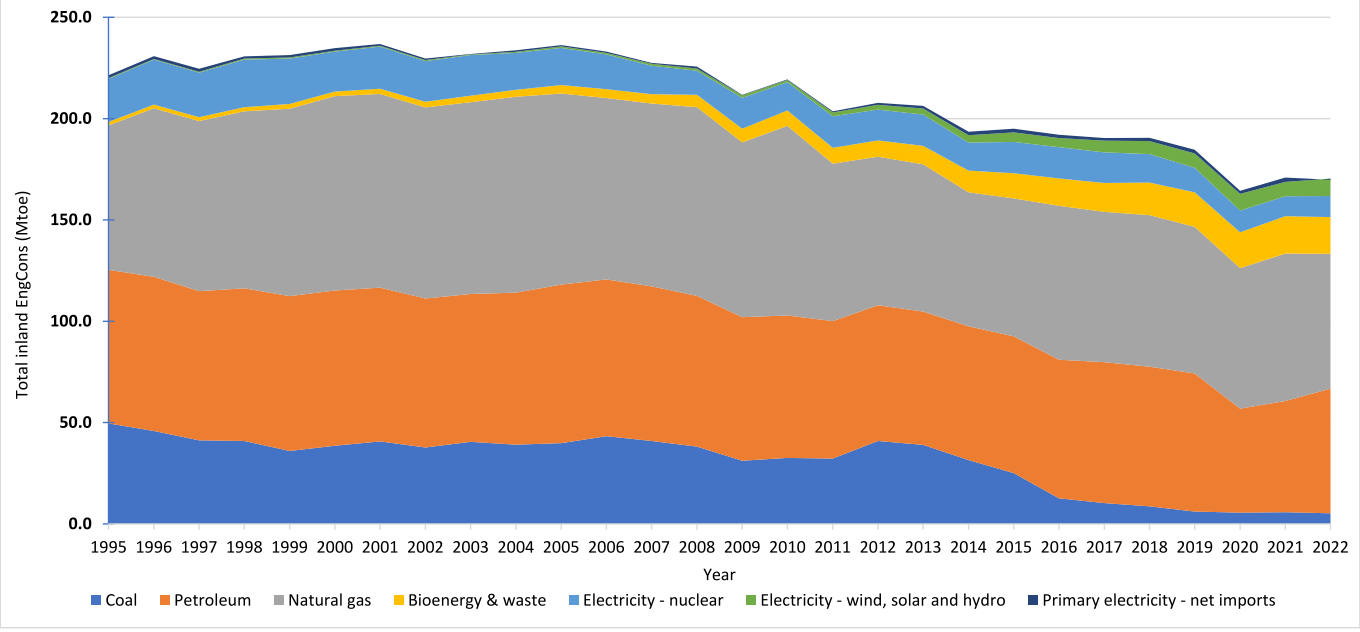


Fig. 1. Annual national inland final EC for primary fuel input basis in million tonnes of oil equivalent (Mtoe) (Sources: UK's Department of Energy Security and Net Zero).

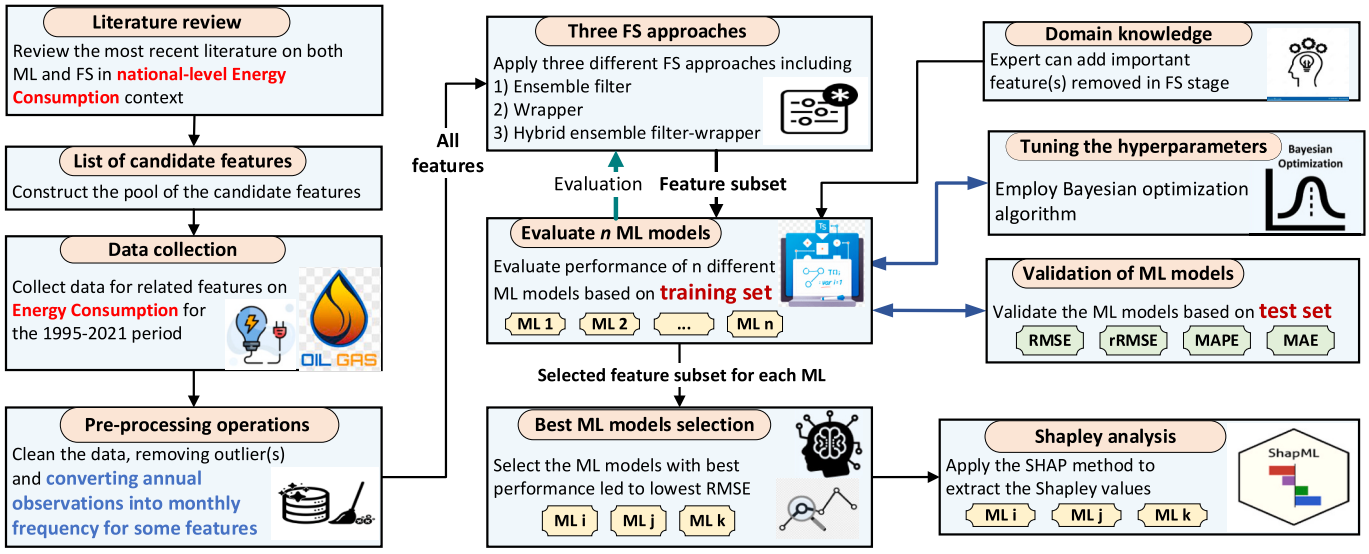


Fig. 2. The proposed framework of the integrated three FS approaches and interpretable ML models for forecasting national inland EC.

models [10]. The incorporation of Shapley additive explanation (SHAP) analysis, following the suggestion by Lundberg and Lee [11], significantly elevates the paradigm. While few prior studies predominantly focused on a single filter/embedded approach individually, our research takes a pioneering step by combining three FS paradigms (ensemble filter, wrapper, and hybrid ensemble filter-wrapper) with ML models. This comprehensive framework enables a unique comparison of these three FS paradigms, making a novel contribution to the field of EC forecasting.

2. Related literature

2.1. FS for energy consumption forecasting

FS approaches, which are commonly divided into three categories – filters, embedding, wrappers methods – serve a significant function in boosting predictive models [12,13]. Filter approaches deploy metrics like mutual information, Euclidean distance, and Pearson correlation to estimate feature relevance based on data properties [14]. While filter approaches excel in computational efficiency, they may not optimize feature sets for specific models [15]. Embedded approaches, meanwhile, generate feature significance directly from model-learned pattern, which are commonly seen in algorithms such as Gradient Boosting or Random Forest [14]. Wrapper approaches, on the other side, integrate FS with an optimization problem that is wrapped around an ML algorithm, resulting in superior precision. However, their computing requirements might make them impractical [16]. Simplified models can minimize computing load while preserving efficacy in specific circumstances. Finally, hybrid approaches integrate many strategies to achieve a balance of computational efficiency and feature quality [17].Understanding the various features selection approaches is critical for developing strong and efficient forecasting models. Meanwhile, given the intricacies and diversity of EC patterns in multiple sectors, coupled with the fact that several ML-based forecasting models likely work best with combinations of key chosen factors, FS remains crucial. Therefore, several studies have employed FS as an initial stage before employing ML techniques to predict their respective focused features in different energy fields. For instance, Qiao et al. [18] employed FS approaches in their EC and CO2 emissions forecasting. In contrast to previous research, which often selected a small number of key features, that study utilized 24 features, including a wide range of socioeconomic, transportation and energy-related features. Similarly, Aras and Hanifi Van [7] investigated a comprehensive comparison of three distinctive FS methods, specifically Lasso, Boruta, and OMP, combined with seven ML techniques in the framework of projecting EC and CO2 emissions. In a separate effort, Hoxha et al. [19] focused on forecasting Transportation energy demand utilizing ML approaches, and they methodically constructed a FS framework to determine the most important inputs. a recently published study by Zyl et al. [17] predicted electric energy demand in Panama combining Gradient-weighted Class Activation Mapping (Grad-CAM) and SHAP as FS methods with convolutional neural network (CNN), covering a wide range of meteorological and temporal factors. This study prominently emphasized the significance of interpretable AI models as an FS approach in EC context. In addition, Kamani and Ardehali [8] forecasted electrical EC, with a focus on solar and wind energy sources. They employed PSO and *E*-PSO algorithms for optimizing hyperparameters. Moreover, Lv and Wang [20] contributed to the field by presenting an efficient model for short-term wind speed prediction that takes into consideration a variety of meteorological factors. To identify relevant meteorological elements, they developed a filter-wrapper technique that used K-medoid clustering. In summary, these studies demonstrated that ML models that used FS approaches frequently outperformed those that did not. The method of systematically selecting significant features assisted substantially to the models' generalizability and precision.

2.2. ML for energy consumption forecasting

The connections between economic growth and EC has been thoroughly studied in the literature, with the premise that energy demand rises in tandem with economic expansion [7]. To that aim, predicting EC is becoming more imperative in order to formulate inclusive policies and therefore, many ML methods for EC forecasting including ANN, MLR, SVM, and RF have been developed in various industries during the last two decades. ML-based forecasting approaches for forecasting EC have demonstrated significant performance, as evidenced in prevailing literature. For instance, Liu and Li [3] predicted UK EC using BP-NN, MLR, and LS-SVM, taking into account factors such as population, GDP, temperature, sunlight, rainfall, and frost days. Ağbulut [5] forecasted vehicle transportation energy demand and CO2E in Turkey using DL, ANN, and SVM algorithms, coupled with GDP and population features. Moreover, Li et al. [9] utilized four ML models for quarterly automobile gasoline demand forecasting in Australia spanning from 1974Q1 to 2019Q2 using autoregressive and structural methods. Wang and Cao [21] forecast EC in China using two grey multi-variable models, and considered EC's interplay with urbanization, and GDP. Whereas,

Kamani and Ardehali [8] employed an ANN model for long-term forecasting of electrical EC considering solar and wind energy sources.

In spite of that, it is determined from the prevailing literature that there is a perceptible scarcity of sufficiently interpretable ML forecasting models for EC, which is indispensable for facilitating well-informed policy decision-making. To date, only very few studies have explicitly addressed interpretable ML models for forecasting EC at country level. For instance, Qiao et al. [18] studied the forecasting of EC and CO2 emissions of transportation sector of the UK using interpretable ML models. Besides, Aras and Hanifi Van [7] also forecasted EC in Turkey utilizing interpretable RF, ANN, KNN, and SHAP models. In summary, prior studies has primarily centred on annual national EC forecasts for all energy sources (except Li et al. [9] used quarterly automobile fuel EC), with relatively limited attention directed toward monthly national EC forecasts, most likely due to the lack of monthly EC data. Moreover, despite the pivotal role of FS in augmenting prediction models, there exists a dearth of studies that systematically compare and evaluate the efficiency of diverse FS approaches, including filter, wrapper, and hybrid filter-wrapper methods. Table 1 provides an overview of studies related to ML-based forecasting models and FS approaches for EC, facilitating comparative analysis. In addressing these gaps within the prevailing pertinent literature, our study emerges as pioneering and well-positioned to advance the field, contributing to more precise, efficient, and interpretable ML models for EC forecasts.

3. Research methodology

3.1. Proposed forecasting framework

Fig. 2 illustrates our proposed conceptual framework for forecasting the national aggregate inland EC from all energy sources. The primary objective of this integrated FS-ML approach is to harness the advantages of different FS approaches in identifying the most influential features, ultimately forming a subset that strikes an optimal balance between ML model accuracy and interpretability, which is well described in Chen et al. [10]. Our proposed method's architecture comprises several key steps:

1) Pre-processing operations: In the initial step, raw data is collected from multiple sources and subjected to preprocessing tasks. These include data consistency correction and integration into a coherent dataset. Furthermore, we convert annual observations of certain features using monotonic functions to monthly data. In this step, preprocessing enables us to create temporal features aimed at capturing the seasonal fluctuations of EC.
2) Correlation analysis of variables: This step involves an in-depth examination of the linear relationships between variables.
3) Ensemble FS method: An ensemble filter and embedded FS method is introduced and employed.
4) Tuning the ML hyperparameters: the Bayesian optimisation is utilized to nearly optimize the hyperparameters of the ML models.
5) Integrated wrapper FS and ML Models: This step involves integrating wrapper FS technique with multiple ML models.
6) Comparative analysis of three FS approaches.
7) Selecting the best integrated FS-ML models: Integrated FS-ML models with the highest accuracy rates are chosen for interpretability analysis using SHAP method.
8) Performing Shapley analysis: SHAP method is carried out to determine the contributions of individual features to the best ML model's output.

Within the framework, we acknowledge that, based on domain knowledge, certain features may be deemed crucial and automatically incorporated into the feature list. The following explanations provide a clearer overview of the proposed method and its multiple steps, with improved clarity and organization.

To enhance comprehension of the implementation of three FS approaches, Fig. 3 elucidates the progression of the pre-processed dataset, denoted as S0, through ensemble filter and/or wrapper FS methods. The three distinct feature subsets obtained are as follows: 1) a feature subset derived solely from the ensemble filter (S1), 2) a feature subset exclusively from the wrapper (S2), and 3) a feature subset comprising both ensemble filter and wrapper selections (hybrid) (S1S2). These subsets are subsequently input into the ML models for performance evaluation. The selection of these three FS approaches is rooted in insights gleaned from an extensive review of relevant literature, particularly FS in Biology and Medicine context, which underscores the drawbacks and limitations of various FS techniques. Typically, ensemble filter and hybrid filter-wrapper FS approaches outperform pure filter and wrapper approaches by mitigating some of their limitations [12,13].

3.2. FS methods

In this section, first we focus on discussing four prominent filter/embedded FS methods: Maximum Relevance and Minimum Redundancy (mRMR), Random Forest, Relief FS, and F-test FS. Then, the proposed ensemble FS is illustrated in Subsection 3.2.5.

3.2.1. Maximum relevance and minimum redundancy (mRMR)

The mRMR algorithm is a FS technique that ranks input features based on their relevance to the output feature while eliminating redundant input features [33]. It utilizes Mutual Information (MI) to measure both the relevance and redundancy of features. MI is defined as:

$$I(X,Y)=\iint p(x,y)log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

In this equation, X and Y represent vectors, $p(x,y)$ is the joint probabilistic density, and $p(x)$ and $p(y)$ are the marginal probabilistic densities. Assuming a feature set S with m features $m\,(x_i, i\in(1,m))$, Max-Relevance aims to find a feature subset with the highest relevance to the output feature y is shown as:

$$maxD(S,y), D=\frac{1}{|S|}\sum_{x_i\in S} I(x_i,y) \tag{2}$$

To identify redundant features, Minimum Redundancy is determined based on possible redundancy within Max-Relevance, which is shown as below:

$$minR(S), R=\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i,,x_j) \tag{3}$$

An incremental search method is then employed to find the optimal solution that satisfies both constraints. Assuming there is already a feature set S_{m-1}, the task is to determine the *mth* feature from the set $\{X-S_{m-1}\}$, which is shown as Eq. (4).

$$max_{x_j\in X-S_{m-1}}\left[I(x_j,y)-\frac{1}{m-1}\sum_{x_i\in S_{m-1}} I(x_i,x_j)\right] \tag{4}$$

3.2.2. Random Forest

Random Forest (RF) is a robust ensemble learning method employed for FS in both classification and regression tasks. It leverages an ensemble of numerous Decision Trees (DTs), with each tree contributing to the selection of the most relevant features during training. These trees operate independently, effectively mitigating the overfitting issues commonly encountered in individual Decision Trees [34]. Notably, Random Forest excels when applied to datasets with diverse characteristics. Within the Random Forest algorithm, FS involves injecting data subsets into each constituent tree. Similar to Decision Tree methodology, Random Forest employs fundamental concepts like entropy Eq. (5) and information gain Eq. (6) to assess feature importance and guide the selection process.

Table 1
Summary of related studies on the application of FS and ML to energy consumption forecast.

Author	Period	Frequency	Country	Prediction ML Models	Input features	Target feature	Feature selection			Feature combination	Interpretable ML
							Filter/ Emb.	Wrapper	Hybrid		
Zyl et al. [17]	2015–2019	Hourly national demand	Panama	CNN	Temperature, humidity, precipitation, wind speed, school holiday, weekend, hour, day, month, year	Electric ED	√	X	X	X	√
Qiao et al. [18]	1990–2019	Annual	UK	SVM, GPR, LSTM, Regression, LSBoost, and MLP	24 socioeconomic, transport- and energy-related features	transport EC and CO2	√	√	X	X	√
Javanmard et al. [22]	1990–2019	Annual	Canada	ARFIMA, MIDAS, SARIMA, AR ARIMA, and SVR	Oil, gas, electricity, and renewable consumptions	ED and CO2	X	X	X	X	X
Nazir et al. [23]	2011–2014	Daily customer's data	London	LSTM, GRN, and TFT	Day, month, year and EC	Electric EC	X	X	X	X	X
Ahmed et al. [24]	2006–2017	Monthly	6 developed & developing countries	LSTM, BiGRU, VMD,	Electric EC	Electric EC	X	X	X	X	X
Rao et al. [25]	1982–2017	Annual	China	SVR-CDSES	Economic development level, population, science and technology, and energy policies	ED	√	X	X	X	X
Saglam et al. [26]	1980–2019	Annual	Turkey	SVM, MNN, WAO	Imports, exports, GDP, and population	ED	X	X	X	X	X
Hoxha et al. [19]	1975–2019	Annual	Turkey	eXtreme Gradient Boosting, Extra Tree Regressor, Random Forest,and DABoost	GDP, year, vehicle miles traveled, population, oil price, passenger miles traveled, and ton-miles traveled	Transport ED	√	X	X	X	X
Kamani and Ardehali [8]	1990–2019	Annual	Developed & Developing countries	ANN and PSO	GDP, Population, Import, Export, and EC	EC	X	X	X	X	X
Atems et al. [27]	1973 M1–2018 M12	Monthly	USA	Recursively identified VAR	Non-renewable energy prices	Renewable electric EC	X	X	X	X	X
Khan and Osińska [28]	1992–2019	Annual	India and Brazil	ONGBM, ONGBM-PSO, and ARIMA	Total EC, Oil, gas, coal, hydroelectricity, nuclear and renewable ECs	EC	X	X	X	X	X
Aras and Hanifi Van [7]	1970–2021	Annual	Turkey	RF, ANN, KNN, GBDT, Adaboost, and SHAP	CO2, GDP, Population, oil consumption, industrial and total EC, Vehicle	EC	√	X	X	X	√
Li et al. [9]	1974Q1-2019Q2	Quarterly	Australia	OLS, Lasso, SVR-linear, and SVR-RBF	Real gasoline price, Australia's real household gross disposable income, population, expenditures on hotels, cafes and restaurants, final consumption expenditure, and purchase of vehicles	Automobile gasoline demand	X	X	X	√	X
Aslan and Beşkirli [29]	1979–2011	Annual	Turkey	AOA	Population, GDP, import and export	ED	X	X	X	X	X

(*continued on next page*)

Table 1 (*continued*)

Author	Period	Frequency	Country	Prediction ML Models	Input features	Target feature	Feature selection			Feature combination	Interpretable ML
							Filter/ Emb.	Wrapper	Hybrid		
Maaouane et al. [30]	1990–2020	Annual	28 European countries	ANN	GDP, population, Pump Price Gasoline, Pump Price Diesel, the Price Index, Purchasing Power Parity, Household Final Consumption Expenditure, and Population Density	EC	X	X	X	X	X
Liu and Li [3]	1993–2019	Annual	UK	BP-NN, MLR, and LS-SVM	Population, GDP, mean temperature, sunshine, rainfall, and frost days	EC	X	X	X	X	X
Ağbulut [5]	1970–2016	Annual	Turkey	DL, ANN, SVM	GDP, Population, Vehicle	Transport ED and CO2	X	X	X	X	X
Wang and Cao [21]	2000–2019	Annual	China	Grey models, SRMGM(1, M) and BRMGM(1, m)	Urbanization, GDP	EC	X	X	X	X	X
Sahraei and Çodur [31]	1990–2019	Annual	Turkey	ANN,ANN-GA, ANN-SA, ANN-PO	GDP, Population, Ton-Km, Vehicle-Km, Passenger-Km, Oil Price	EC	X	X	X	√	X
Zheng et al. [32]	2010–2018	Annual	China	NGBM and PSO	hydroelectricity consumption	Hydroelectricity consumption	X	X	X	X	X

Note: √ for Yes, and X for No. EC-Energy Consumption, ED-Energy demand, CO2-Carbon dioxide emissions, GDP-Gross Domestic Product, ANN- Artificial neural network, LSTM-the Long Short-Term Memory, SVM-Support vector machine, GPR-Gaussian process regression, MLP-Multi-layer perceptron, TFT- Temporal Fusion Transformer, ARIMA-Autoregressive Integrated Moving Average, KNN-K Nearest Neighbors, ARFIMA-Autoregressive Fractional Integrated Moving Average, BiGRU-Bidirectional gated recurrent unit, VMD-Variational mode decomposition, SVR-Support Vector Regression, LASSO-Least Absolute Shrinkage and Selection Operator, RF-Random Forest, MAPE-Mean Absolute Percentage Error, RMSE-Root Mean Squared Error, NRMSE-Normalized Root Mean Square Error, MPE-Mean Percentage Error, RNN-Recurrent Neural Network, AOA-Arithmetic optimization algorithm, GARCH- Generalized Autoregressive Conditional Heteroscedasticity, MIDAS-Mixed Data Sampling, BP-NN- Back Propagation Neural Network, MLR- Multiple Linear Regression, LS-SVM- Least Square Support Vector Machine, GRNN-General Regression Neural Network, EWT-LSTM-Empirical Wavelet Transform, GRN-Gated Residual Network, ONGBM- optimized nonlinear grey Bernoulli model, and ONGBM-PSO- nonlinear grey Bernoulli model with particle swarm optimization, ANN-GA-Genetic Algorithm, ANN-Simulated Annealing (ANN-SA), and Particle Swarm Optimization (ANN-PSO), MNN-Medium Neural Networks, WAO-Whale Optimization Algorithm.

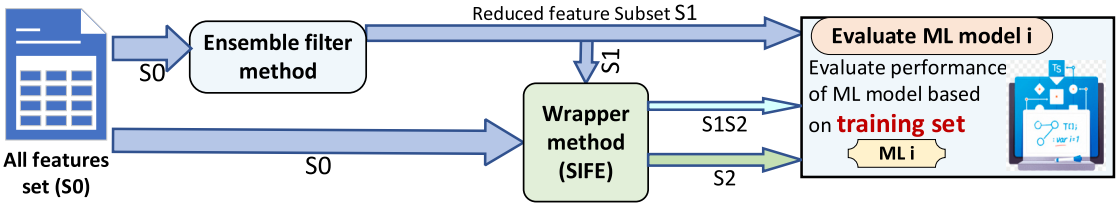


Fig. 3. The schematic diagram on how three FS approaches are integrated with ML models.

$$Entropy(p_1, p_2, \ldots, p_k) = -\sum p_1 log_2(p_1) \tag{5}$$

$$I_G = 1 - \sum_{J=1}^{C} P_J^2 \tag{6}$$

3.2.3. Relief FS

Kira and Rendell [35] introduced the Relief algorithm, inspired by instance-based learning [36]. Relief is a method for individual feature assessment, estimating 'quality' in predicting the endpoint value through feature weights $W[A]$, informally referred to as 'scores' ranging from -1 to $+1$. Initially designed for binary classification, the algorithm lacked a mechanism for missing data, which is addressed in subsequent works referenced in the review section.

The Relief algorithm iteratively processes random training instances R_i, with 'm' as a user-defined parameter. Within each iteration, a 'target' instance R_i is selected, and feature scores in vector W are updated based on observed differences between the target instance and neighboring instances. Distances between the target instance and others are calculated, identifying the two nearest neighbors: one from the same class ($nearest\ hit, H$) and the other from the opposite class ($nearest\ miss, M$). Feature weights are updated based on differences between the target instance and H or M. The *diff* function quantifies the disparity in feature A between instances I_1 and I_2, based on whether A is discrete or continuous.

For discrete features, *diff* is determined by Eq. (7), ensuring that weight updates fall within the $[0, 1]$ range. Final weights are normalized within $[-1, 1]$ by dividing the *diff* output by m. For continuous features, *diff* is defined by Eq. (8). The *diff* function is also used in calculating distances between instances to identify nearest neighbors.

$$diff(A, I_1, I_2) = \begin{cases} 0 \ if \ value\ (A, I_1) = value\ (A, I_2) \\ 1 \ otherwise \end{cases} \tag{7}$$

$$diff(A, I_1, I_2) = \frac{|value\ (A, I_1) - value\ (A, I_2)|}{max(A) - min(A)} \quad (8)$$

3.2.4. F-test FS

The F-test FS is a technique used to rank input features based on their significance for a specific ML task. It employs statistical analysis to determine the relevance of individual features and their impact on the output feature. The F-statistic (F) is utilized to quantify the significance of features. The F-value reflects a feature's significance in explaining the variability of the output feature. FS through the F-test entails three steps: calculating F-values for each feature, ranking these features in descending order based on their F-values, and selecting the top-k features with the highest F-values, where *k* can be a user-defined parameter or determined through cross-validation.

By using the F-test FS, irrelevant or less important features can be identified and removed, leading to a more efficient and accurate ML model. This technique is particularly useful when dealing with high-dimensional datasets where FS can improve model performance and reduce overfitting.

3.2.5. Proposed ensemble FS

In the domain of FS, a significant challenge lies in the selection of the most appropriate filter and embedded FS method(s) for a given dataset [37]. This challenge arises from the inherent diversity among FS techniques, each employing its distinctive logic rooted in statistical measures for assessing feature importance. Consequently, this diversity can yield distinct sets of selected features. To put it simply, a feature considered highly significant in one method may not hold the same level of importance in another. Nevertheless, it's crucial to acknowledge that both filter and embedded methods come with inherent limitations [12]: filter methods may inadvertently overlook interdependencies among input features, while embedded methods rely significantly on ML models.

Ensemble FS, a technique in the realm of ML, is employed to enhance the accuracy and robustness of the FS process [38]. This approach works by amalgamating the outcomes of multiple FS methods to make more informed decisions regarding which features to retain and which to discard. Ensemble FS offers advantages by leveraging the strengths of multiple FS methods and mitigating their individual weaknesses. It aids in reducing the risk of overfitting and enhances the likelihood of selecting the most relevant features tailored to a specific prediction task [38]. The desire for an ensemble FS method stems from its potential to enhance both the stability and interpretability of feature selection, thereby facilitating the subsequent process of EC prediction [39,40].

Let us delve into the procedure of our proposed ensemble FS method as depicted in Fig. 4:

1) Ensemble FS parameter determination: Begin by considering all N candidate features and specifying the desired number of features (m) from K base FS methods, along with establishing the minimum number of votes (v) required from these base FS methods.
2) Selection of base FS methods: Initiate the process by selecting a combination of base FS methods, encompassing both filter and embedded FS techniques.
3) Feature scoring and ranking: Each chosen base FS method calculates importance scores for all features. Through an iterative process, the feature with the least importance score is eliminated until the desired number of features is achieved. The resulting list includes the 'm' most important features along with their respective importance scores.

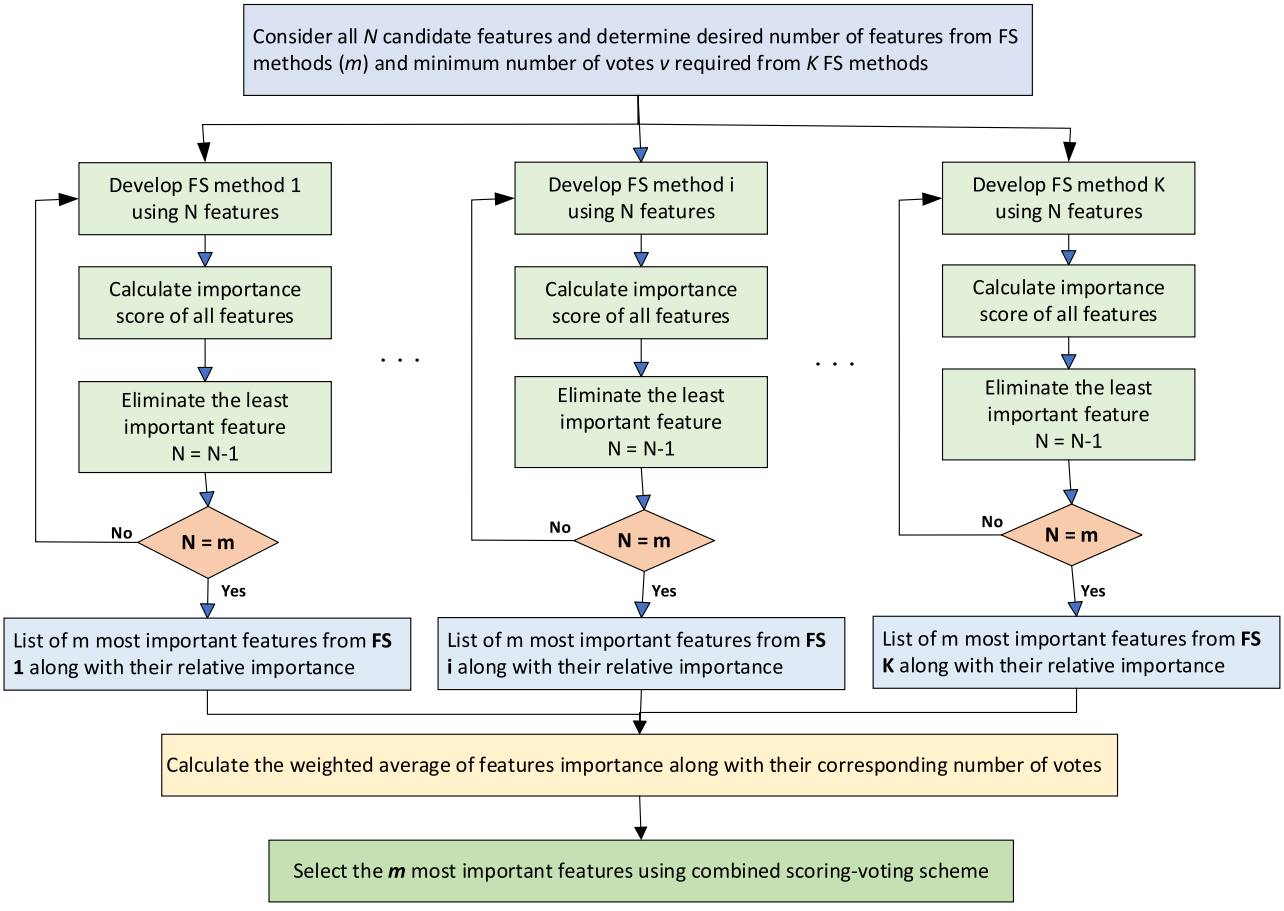


Fig. 4. The conceptual framework of the proposed ensemble FS.

4) Aggregating feature importance: Combine feature importance scores, assigning weights based on each base FS method. Employ a weighted averaging approach coupled with a voting mechanism to create a unified ranking reflecting the average importance of features.
5) Final FS: Retain the 'm' most important features, identified by the highest importance scores and voted by at least two base FS methods. Exclude the remaining features from consideration.

In our envisioned ensemble FS approach, we integrate four distinct base FS methods: two multivariate filter methods (mRMR and Relief), one univariate filter method (F-test), and one embedded method (Random Forest). The objective is to identify the most crucial features by combining different FS categories. Filter methods, renowned for their computational efficiency and independence from ML models, are less susceptible to overfitting but do not directly interact with ML models [12,37]. Unlike univariate filters, multivariate filters have the advantage of considering feature redundancy among exploratory features.

The primary reasons for selecting these four FS methods are: 1) They offer a reasonable combination of different FS categories, with an emphasis on multivariate filters, 2) They are among the most widely used FS methods in the literature, and 3) Their implementations are available in MATLAB. For a deeper understanding of the strengths, weaknesses, and illustrative examples of the main feature selection categories, readers are encouraged to consult the works of Pudjihartono et al. [12] and Theng and Bhoyar [13].

3.3. ML models

In this academic section, we focus on discussing four prominent ML models, which include Support Vector Machine (SVM) with two variations (the RBF kernel and the Linear kernel), Gaussian Process Regression (GPR), LASSO, and Gradient Tree Boosting with Least Squares (LSBoost). Based on a comprehensive review of the literature, we have identified these diverse ML models as effective tools for forecasting energy consumption. Nevertheless, neural network-based models were omitted due to their computational intensity, rendering them incompatible with Wrapper FS, particularly when dealing with ML models involving thousands of feature combinations [41].

3.3.1. Support vector machine (SVM)

Support Vector Machines (SVMs) are a versatile ML approach capable of handling regression and classification tasks with high precision, thanks to their robust mathematical underpinnings. SVMs operate by systematically positioning data points in an n-dimensional space, where n represents the number of attributes in each data sample, treating attribute values as coordinates [39,40]. The primary goal of the SVM algorithm is to establish a decision boundary that effectively separates and categorizes these data points. This boundary's selection is pivotal for the SVM's performance.

During training, SVM aims to identify a decision boundary that maximizes the margin between data points from different classes, enhancing predictive capabilities. When y signifies target classes, SVM constructs a hyperplane H. Determining this hyperplane involves complex mathematical procedures based on input data x. The weight vector a holds feature weightings, b represents a bias term, and $\varnothing(0)$ signifies a fixed feature-space transformation, collectively influencing the decision boundary, which is shown as below:

$$H : a^T \varnothing(x) + b = 0 \tag{9}$$

Additionally, when applying Support Vector Regression (SVR) to the problem of regression, SVMs can be extended with different kernels, such as the Radial Basis Function (RBF) kernel and the Linear kernel. SVR-RBF leverages the non-linear characteristics of the RBF kernel to capture complex relationships between data points, while SVR-Linear employs a linear kernel for capturing linear relationships. These kernels play a vital role in tailoring SVMs for regression tasks, making them versatile tools for various types of data analysis and prediction.

3.3.2. Gaussian process regression (GPR)

One key feature of GPR is its ability to make predictions while quantifying the associated uncertainty [9]. The algorithm relies on a training set, represented as $D = (X, y) = \{(X_i, y_i) | i = 1, \ldots, N\}$, where X is the input vector, and y is the output or target feature. Given a novel input X^*, GPR predicts the corresponding output $\hat{y}^*$ using Eq. (10).

$$\hat{y}^* = K(X^*, X)K(X, X)^{-1}y \tag{10}$$

This equation defines the core predictive capabilities of GPR. The derivation process involves Eqs. (11) and (12), with Eq. (11) introducing the initial assumption and Eq. (12) derived from the characteristics of the conditional distribution of a multidimensional Gaussian distribution [42].

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}\right) \tag{11}$$

$$y^* | y \sim \mathcal{N}\left(K(X^*,X)K(X,X)^{-1}y, K(X^*,X^*) - K(X^*,X)K(X,X)^{-1}K(X,X^*)\right) \tag{12}$$

To maximize $p(\hat{y}^* | y)$, the prediction $\hat{y}^* = K(X^*, X)K(X, X)^{-1}y$.

3.3.3. LASSO

Linear regression (LR) is a statistical method used to examine the relationship between a designated dependent feature and a given set of independent features, as outlined by Centofanti et al. [43]. In this context, it is pertinent to recognize the existence of m independent features. Within the regression equation, β_0 represents the constant term, and β_1 through β_m denote the coefficients associated with these independent features. Additionally, ε is introduced as the stochastic error term.

It is noteworthy that the m^{th} regression coefficient, β_m, signifies the anticipated change in the dependent feature Y for each unit alteration in the m^{th} independent feature x_m. This interpretation assumes that the expected value of the error term, $E(\varepsilon)$, is zero, and it can be expressed mathematically as $\beta_m = \frac{\partial E(Y)}{\partial X_m}$. Also, Eq. (13) represents the LR model.

$$Y = \beta_0 + cX_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon \tag{13}$$

In the context of regression analysis, an alternative method called LASSO (Least Absolute Shrinkage and Selection Operator) is utilized. LASSO, unlike traditional linear regression, incorporates a penalty term that encourages sparsity in the coefficient estimates [44]. This means that LASSO not only identifies the relationship between dependent and independent features but also performs FS by shrinking some coefficients to zero. This can be particularly useful when dealing with high-dimensional data, as it helps in identifying the most influential independent features while discarding less important ones. Eq. (14) represents the LASSO model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon \tag{14}$$

LASSO is a valuable technique in regression analysis, especially when seeking a simpler and more interpretable model with a subset of the most relevant features.

3.3.4. Gradient tree boosting with least squares (LSBoost)

LSBoost, a meta-learning technique, is a methodology that amalgamates a specified count of weak tree-learners to enhance predictive performance [45]. This approach unfolds with a sequential training of these weak learners in the form of decision trees, followed by the refinement of their predictions through fitting the residual errors. The criterion for loss in the LSBoost framework centers on the least squares.

Considering a training set $\{(x_i, y_i)\}_{i=1}^{n}$ for $i = 1$ to n, the loss function $L(y, F)$ is defined as $\frac{(y-F)^2}{2}$, and a regression function $F_m(x)$ is introduced for each iteration, where m signifies the iteration number. At each iteration, the formulation progresses as:

$$\widetilde{y_i} = y_i - F_{m-1}(x_i), \text{for i} = 1, 2, \ldots, \text{N} \tag{15}$$

The parameters (ρ_m, α_m) are derived by minimizing the sum of squared differences between $\widetilde{y_i}$_approximated and $\rho h(x_i; \alpha)$, as expressed in Eq. (16).

Subsequently, the regression function is updated as $F_m(x) = F_{m-1}(x) + \rho_m h((\text{x}; \alpha_m)$, where $h(x_i; \alpha)$ represents a parameterized function characterized by the parameter α_m, and ρ_m signifies the incremental step in the LSBoost process.

$$(\rho_m, \alpha_m) = argmin_{\rho,\alpha} \sum_{i=1}^{N} [\widetilde{y_i} - \rho h(x_i; \alpha)]^2 \tag{16}$$

3.4. The Bayesian optimization algorithm

The Bayesian optimization algorithm stands out as a global optimization method tailored for handling costly objective functions. Unlike traditional approaches, it is free from population and genetic operators (mutation, crossover, and selection). This algorithm employs a Gaussian process to calculate an acquisition function, effectively evaluating the objective function [46]. Additionally, Bayesian optimization leverages its historical evolution to enhance its performance, utilizing previously acquired statistics to guide the search for optimal solutions. Literature suggests that Bayesian optimization surpasses the effectiveness of grid search and random search, showcasing competitiveness with emerging evolutionary optimization algorithms. [46,47].

3.5. Wrapper FS

The Set-based Integer-coded Fuzzy granular Evolutionary (SIFE), an evolutionary algorithm widely recognized for its efficiency, is designed to handle FS in both high-dimensional and low-dimensional scenarios [48]. To improve its search strategy, SIFE uses a unique three-parent crossover method rooted in set-theory principles like 'union' and 'intersection'. It also incorporates fuzzy granulation, which aids in tasks like initializing the population and selecting elite individuals. This incorporation encourages diversity within the population over successive generations and reduces the need for additional fitness evaluations. The primary goal of this approach is to find a balanced trade-off between exploring new solutions and exploiting known ones while keeping computational complexity in check. SIFE is chosen for this research due to its proven ability to navigate and optimize search spaces of varying dimensions effectively.

SIFE incorporates a systematic approach in each generation, evaluating solution rankings for selection and reproduction. A pivotal objective in FS is optimizing a quality measure that balances two primary goals within machine learning (ML) algorithms. This entails minimizing the ML algorithm's error metric while identifying a compact subset of highly relevant, minimally redundant features. To achieve this, the SIFE framework introduces an objective function:

$$minimize\ F(s_i) = w_1 \times Er(s_i) + w_2 \times Ps\ (s_i) \forall s_i \in \Omega, \tag{17}$$

addressing both objectives concurrently.

Feature subset and is defined as:

This function evaluates candidate feature subset s_i from the set of possible features Ω, considering the error metric Er such as RSME or MAPE and the percentage of selected features Ps. Here, w_1 and w_2 define the relative significance of these objectives, set to 0.80 and 0.20 respectively, owing to the study's limited feature count.

3.6. ML interpretation by SHAP method

Critique of black-box ML models due to their limited interpretability has led to a demand for quantitative analysis of the relationship between input and target features in decision-making. Shapley Additive Explanations (SHAP) provides a robust method for interpreting these models. SHAP utilizes the classical Shapley value from game theory to link optimal credit allocation and local explanations. This is achieved by breaking down the model's predictions into the sum of individual feature impacts, improving our understanding of feature importance and aiding informed decision-making. To compute SHAP values, a linear explanation model is used as an interpretable approximation of the ML model, which is shown as:

$$g(z^{'}) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i^{'} \tag{18}$$

Here, $z^{'} \in \{0, 1\}^M$ denotes whether a feature contributes to estimating the output feature, M represents the number of input features, ϕ_i signifies the SHAP value of the $i - th$ feature, and ϕ_0 represents the mean value of the output feature [49]. The SHAP value evaluates feature importance by comparing the model's prediction performance with and without each feature in various feature combinations, as shown:

$$\phi_i = \sum_{S \subseteq z^{'} \{i\}} \frac{|S|!(M - |S| - 1)}{M!} [f_x(S \cup \{i\}) - f_x(S)] \tag{19}$$

In this equation, S signifies the set of non-zero $z^{'}$, and $f_x(S) = E[|f(x)x_S|]$ represents the expected model outcome of $f(x)$ influenced by S.

3.7. Model evaluation

In order to comprehend the performance of the model concerning the national EC within the UK, we have included a set of evaluation metrics as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2} \tag{20}$$

$$rRMSE = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2}{\sum_{i=1}^{n} (p_i)^2}} \tag{21}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - p_i| \tag{22}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - p_i}{y_i} \right| \tag{23}$$

$$R^2 = 1 - \frac{\left(\sum_{i=1}^{n} (y_i - p_i)^2 \right)}{\left(\sum_{i=1}^{n} (y_i - \overline{y_i})^2 \right)} \tag{24}$$

In the given equations, y_i represents the actual or observed values of the variable being predicted or modeled, p_i symbolizes the predicted values generated by the model, and $\overline{y_i}$ usually denotes the mean (average) of the observed or actual values y_i It represents the average value of the data points.

4. Experimental setting

This research study employs MATLAB R2022b for the implementation of all experiments including ensemble FS, ML models, wrapper, and hybrid FS. MATLAB R2022b is utilized for SHAP analysis generating time series plot and heatmap, and Excel for Microsoft 365 is used for

producing spider plots.

4.1. Data description

This study centers around national EC within all sectors of the UK, covering the time period from 1995 to 2021. Notably, the years 2020 and 2021, largely affected by the disruptive forces of the Covid-19 Pandemic, have been included in this analysis. It's worth highlighting that the monthly EC values for these two years exhibit some deviation from corresponding values in previous years, which subsequently results in higher errors when applied to ML models. An interesting observation emerges: when the years 2020 and 2021 are omitted from the experimental dataset, the performance of the ML models presents slight improvements.

The dataset depicting seasonally adjusted total EC and unadjusted total EC at an annualized rate in the UK for the period from January 1995 to March 2023 is visually presented in Fig. 5. The chronological trajectory of inland final EC exhibits discernible patterns. Specifically, an initial phase of growth is succeeded by a significant and sustained decline after 2007–08 global financial crisis, a trend that persists until 2023.

After an extensive review of the relevant literature, this study integrates 18 distinct features (variables) to unravel the intricate interplay among energy consumption, economic performance, and environmental factors. In Table 2, you will find the comprehensive list of these 18 features along with their respective abbreviations.

The target feature of our analysis, 'total final EC,' serves as a representation of the cumulative energy consumption derived from all energy sources across residential, transportation, commercial, and industrial sectors. This, in essence, reflects a foundational indicator of national energy demand within our study.

Among the primary explanatory features are the consumptions of various primary fuels, including Coal, Petroleum, Natural gas, Bioenergy & waste, Electricity-nuclear, Electricity-wind, solar & hydro, and Share of renewable energy. These features, as highlighted by Javanmard et al. [22] and Khan and Osińska [28], provide invaluable insights into the utilization of specific energy sources, each offering unique perspectives on consumption patterns. GDP and Population stand out as two key explanatory features across numerous studies (as shown in Table 1). GDP per capita serves as a crucial indicator of economic well-being, while Population represents the overall demographic size, influencing energy consumption scaling.

Table 2
Eighteen features and their corresponding abbreviation.

Feature	Abbreviation	Feature	Abbreviation
Total monthly final EC	EC	Temperature	Temp
Coal	Coal	Population	Pop
Petroleum	Pet	Energy intensity per unit of GDP	EI
Natural gas	Gas	Real price - Gas	P-Gas
Bioenergy & waste	B&W	Real price- Electricity	P-El
Electricity - nuclear	El-Nu	Real price - Motor fuel & oil	P-Fuel
Electricity – wind, solar & hydro	El-WSH	Sine month	SinM
Share of renewable energy	SRnw	Cosine month	CosM
GDP per capita	GDP	Month number	Mon#

Temperature, an environmental feature, plays a pivotal role, significantly impacting national-level energy consumption patterns, as noted by Zyl et al. [17] and Liu and Li [3]. Energy intensity serves to quantify energy consumption relative to economic output, providing essential context for understanding energy usage trends. To our knowledge, no study has considered Energy intensity as a explanatory feature. Moreover, real prices of gas, electricity, and motor fuel, along with oil, are considered to encompass the pricing dynamics of respective energy sources, as discussed by Atems et al. [27] and Maaouane et al. [30]. These factors collectively contribute to a comprehensive understanding of energy consumption dynamics and their interplay with economic, demographic, environmental, and pricing factors.

Among the 18 features, 3 features have been designated as temporal features. Month number (Mon#), Sine month (SinM), and Cosine month (CosM) are pivotal for capturing temporal and cyclical fluctuations within the dataset [50,51]. The month number serves as a fundamental metric for encapsulating inherent seasonality. While the month number feature enriches the analytical depth of the study by facilitating the identification of patterns and temporal shifts, its standalone usage may pose constraints, particularly in time series analysis [50,51]. Hence, complementary features like Sine month and Cosine month are integrated to augment the temporal analysis, offering a more comprehensive

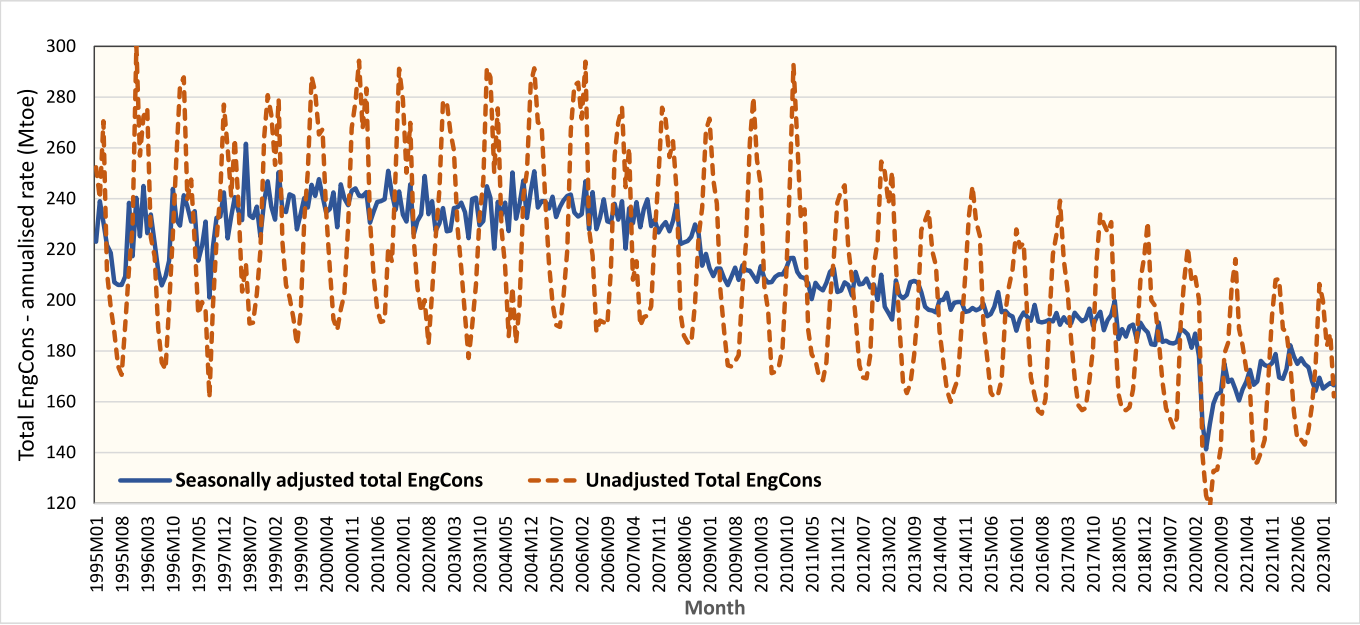


Fig. 5. Aggregate inland final EC at an annualized rate in the UK for the period from Jan 1995 to Mar 2023.

perspective on the dataset's dynamics. The descriptive statistics of the 18 features considered for UK's national EC is presented in Appendix Table A1.

The data was collected from the multiple sources: UK Department for Energy Security and Net Zero, UK Office for National Statistics, and International Energy Agency (World Energy Balances Highlights).

Through a combination of exploratory data analysis and domain knowledge, it becomes evident that certain features, such as the real price of solid fuels, liquid fuels, and domestic fuels, as well as meteorological variables like average wind speed, humidity, rainfall, and sunshine, along with socioeconomic factors like the unemployment rate, need not be included in the list of candidate primary features. The strategic exclusion of these features from consideration helps maintain a manageable size for the list of candidate features. This deliberate pruning is instrumental in addressing challenges associated with multicollinearity, and it concurrently aids in reducing the feature space for the wrapper FS process.

4.2. Data preprocessing

For preprocessing, the study utilized cyclical feature encoding and ordinal encoding, specifically focusing on time representation, notably the month. While ordinal encoding lacks accounting for cyclic categorical features like time data, cyclical feature encoding resolved this through sine and cosine transformations of the cyclic feature x (*month*).

Data normalization, a critical step to map data onto a unit sphere, addressed variations in feature dimension scales. The normalization process, as represented by $x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$, normalizes data points x between minimum (x_{min}) and maximum (x_{max}) values to enhance ML algorithm performance reliant on Euclidean distance metrics.

Furthermore, it is noteworthy that only 2 out of the 18 features, specifically Population and Energy intensity, exhibit monotonic functions, as depicted in Fig. 6, and are available on an annual basis. To obtain monthly observations for these variables, we utilize Chow-Lin Regression-based interpolation, a method employed for converting low-frequency data to high-frequency data, implemented in EViews software [52]. However, it is crucial to emphasize that converting variables with annual observations, such as Temperature and Gas, which experience seasonal fluctuations, or those with uncertain changes, such as primary fuels' price and GDP, to low frequency is not appropriate. Such conversions can lead to misleading interpretations and inaccurate analyses.

4.3. Hyperparameter setting of MLs

Hyperparameter optimization constitutes a critical facet in the realm of ML, as it holds the key to unlocking the full potential of predictive models and engendering optimal performance. The endeavor to systematically and judiciously search for the most suitable hyperparameter settings is a paramount undertaking in ML research. In this study, we partitioned 80% of the data for training and hyperparameter tuning, reserving the remaining 20% as the test set. To tune hyperparameters, Bayesian Optimization was employed using MATLAB R2023b's Statistics and Machine Learning toolbox. It's noteworthy that our experiments revealed limited effectiveness of hyperparameter tuning, with only slight performance improvements observed in a few out of 15 configurations. The results of the hyperparameter optimization process using Bayesian optimization are presented in Table 3.

Table 4 displays the general parameter settings for SIFE, with a population size of 120 and a maximum iteration of 60.

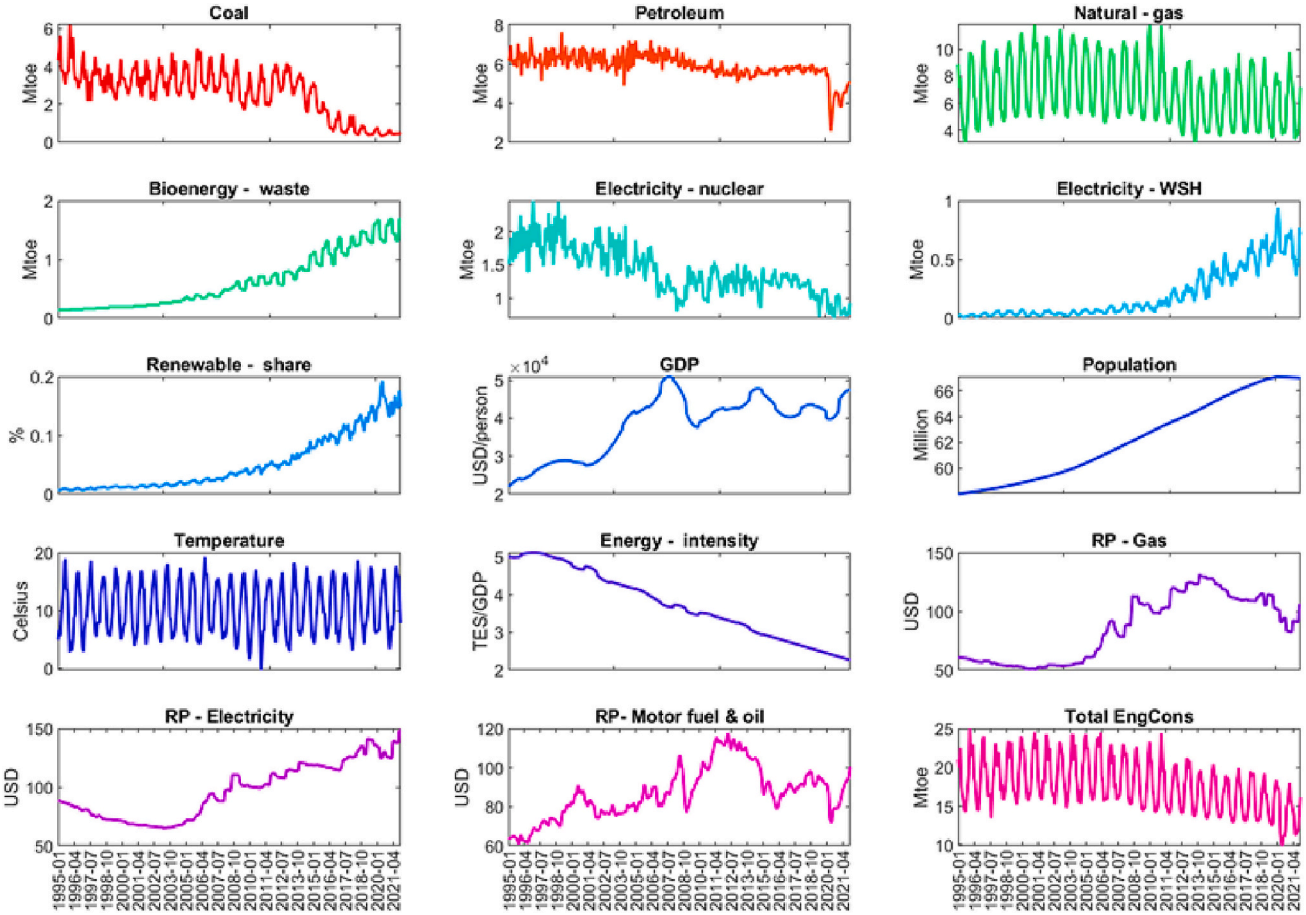


Fig. 6. Visualizing the time series plot of 15 numerical features of the dataset.

Table 3
The summary of hyperparameters for five ML models.

ML Model	Hyperparameter	Value/Type
LASSO	Learner	Least Square
	Regularization	Lasso
	lambda	1.16e-04
SVR-Linear	Box constraint	10
	Epsilon	0.005
SVR-RBF	Box constraint I	10.5
	Gamma (γ)	35
	Epsilon	0.003
LSBoost	Maximum number splits	12
	Pruning	Off
	Number of learning cycles	8
	Learning rate	0.28
GPR	Sigma	0.046
	Prediction method	Subset of regressors approximation
	Basis Function	Peru Quadratic

Table 4
SIFE's parameters.

Parameter	Value
Population Size	120
Maximum iteration	60
Crossover probability	0.6
Mutation probability	0.5

5. Results and analysis

5.1. Primary visual analysis

In this section, we conducted a thorough analysis of each primary feature. Fig. 6 presents a visual representation of the time series patterns characterizing these features. Moreover, taking a broad view of GDP per capita reveals a notable and steady increase, indicating strong economic growth. Simultaneously, there is an upward trend in the use of Bioenergy & waste, the Share of renewable energy, Petroleum, and Electricity from wind, solar, and hydro sources. In contrast, there is a noticeable decrease in the use of Coal, Energy intensity, and Nuclear-electricity. Considering the observed fluctuations and shifts within these features, it is essential to conduct a comprehensive analysis to understand their underlying causes. The rise in energy prices, as mentioned earlier, is a significant factor contributing to these changes. Higher energy prices play a crucial role in influencing consumer choices, encouraging a shift toward more cost-efficient and sustainable energy sources. This transition aligns with environmental and economic considerations, supporting the increased use of Bioenergy & waste, the Share of renewable energy, and Electricity from wind, solar, and hydro sources.

Furthermore, the decline in the use of Coal and Nuclear-electricity is likely driven by environmental concerns and the increasing competitiveness of renewable energy sources. Notably, the observed price fluctuations, especially in Real price - Gas, Real price - Electricity, and Real price - Motor fuel & oil, are closely tied to global market dynamics, supply and demand dynamics, and geopolitical influences. The rising GDP per capita may lead to increased energy consumption, especially in sectors closely linked to economic growth. It is evident that the fluctuations in Coal, Gas, Petroleum, Electricity (nuclear and WSH) and share of renewable are due to the seasonal temperature changes.

5.2. Pearson correlation analysis

The fluctuations and transitions witnessed within the features under investigation are the result of intricate interactions among a medley of economic, environmental, and market-driven factors. These discernible alterations underscore the intricate interplay that characterizes the interrelationships between energy consumption and other pertinent features. As such, it is incumbent upon us to embark on a comprehensive analysis that embraces this multifaceted landscape of influences. In this section of our study, we endeavor to unravel the complexities and dynamics inherent to these features by employing heatmap of Pearson correlation analysis, which is visually represented in Fig. 7.

Given that all 15 input primary variables are continuous numerical data, we conducted a comprehensive analysis using Pearson correlation coefficients with a two-tailed test. The results reveal several noteworthy findings. Firstly, certain correlation values emerged as statistically significant at a confidence level of $\alpha = 0.01$. Notably, some correlations exhibited absolute values surpassing a threshold of 0.90. One striking observation is the nearly perfect negative correlation between population and energy intensity, denoted by $r(322) = -0.99$. Additionally, the variables Electricity-WSH, Renewable-share, and Population exhibit an exceptionally strong positive correlation with Bioenergy & Waste, reflected in their correlation coefficient of $r(28) = 0.96$.

Moreover, in the correlation analysis of the target variable, it is observed that Natural gas and Coal demonstrate remarkably positive correlations with Total EC(t), with coefficients of $r(322) = 0.94$ and 0.76, respectively. Interestingly, Temperature exhibits a robust negative correlation of -0.82 with Total EC(t). Upon closer examination of Fig. 6, we posit that the (very) strong correlation between Natural gas, Temperature, and Coal with Total EC(t) can be attributed to corresponding fluctuations during different seasons. Specifically, the consumptions of Natural gas, coal, and Total EC peak during colder months and decline during warmer months, resulting in these pronounced correlations.

Of particular note is the moderately negative correlation of - 0.49 between Population and Total EC(t), implying that, despite a consistent increase in population, Total EC in the UK is on a declining trend. This intriguing finding prompts further exploration into the dynamics influencing EC patterns in the context of population growth.

5.3. FS results

In this section, we scrutinize the outcomes of the FS algorithms, focusing on the identification of the ten most important features for inclusion in our input feature subset. The comprehensive results of the FS algorithms are presented in Table 5. Table 5 encompasses a compilation of all the primary features and temporal features employed throughout the study. It further includes a column titled "Total count," which signifies the number of features designated as significant by these algorithms. Specifically, dedicated rows are allocated to the mRMR, RF, F-test, and Relief algorithms, elucidating the relative importance of features achieved by each FS method.

For instance, within the mRMR algorithm, the feature EC(t) attains a relative importance of 1, signifying its preeminent significance within this particular algorithm. Following EC(t), the features P-Gas and CosM occupy the second and third positions in terms of importance, boasting respective importance values of 0.804 and 0.774 in the mRMR algorithm. Additionally, this algorithm identifies features such as El-Nuc, GDP, EI, P-El, P-Fuel, SinM, and Mn# as less significant features, underscoring their reduced impact.

The "Vote" row assigned to each feature serves to accentuate their significance by revealing their collective importance across four filter FS methods, specifically, mRMR, RF, F-test, and Relief. Furthermore, the "Average importance" row computes the average importance score of features across the various methods. For instance, the EC(t) feature registers a value of 1 in both the mRMR and F-test algorithms, while in the RF and Relief methods, it garners values of 0.740 and 0.412, respectively. The resultant average importance score is 0.788. This amalgamation of importance scores forms the basis for the final feature ranking, encapsulated in the "Final rank" row.

In accordance with our refined methodology, the culminating assessment of feature importance based on average scores now yields a ranking of features. Notably, CosM, EC(t), and Gas emerge as the top

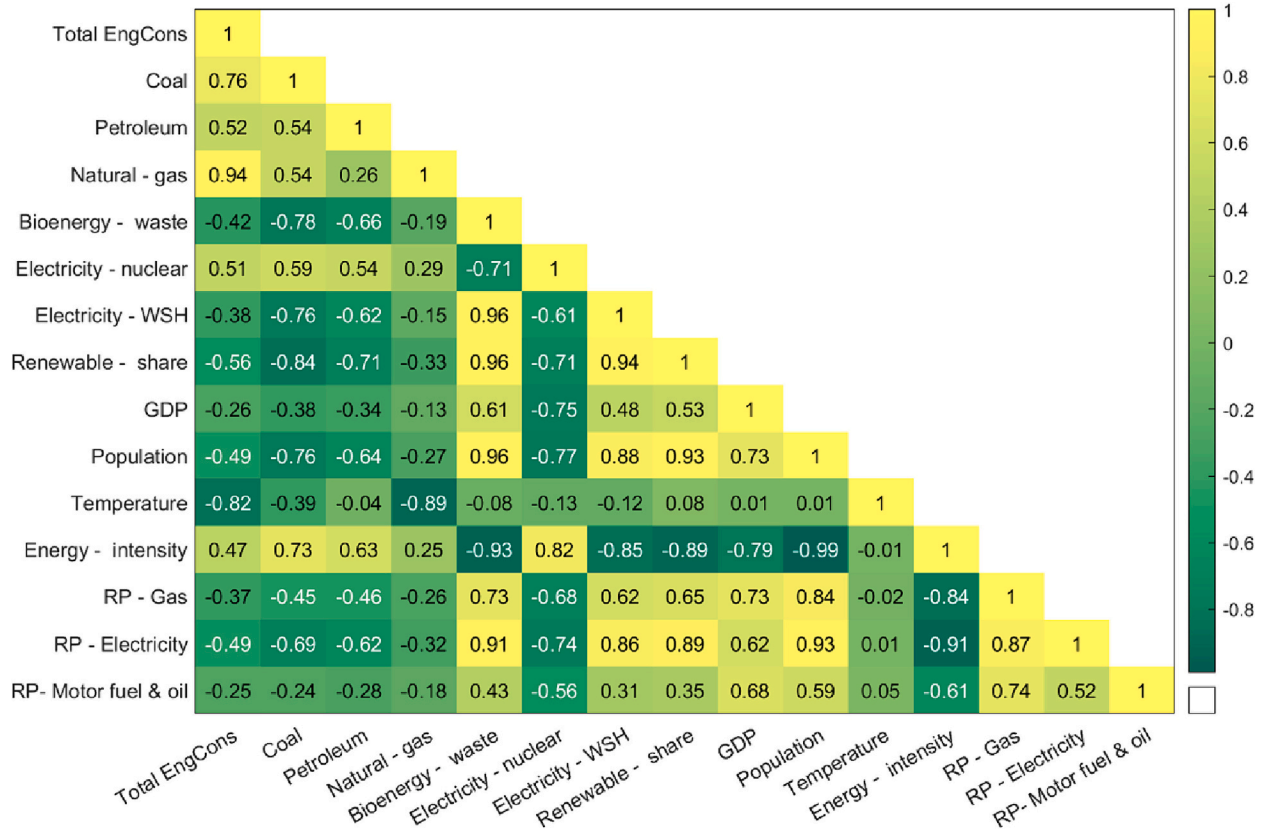


Fig. 7. The heatmap of Pearson correlation coefficient for EC(t).

three features in terms of significance, securing the first, second, and third positions, respectively, across all four mRMR, RF, F-test, and Relief algorithms. Energy intensity (EI) emerges as the third most significant feature according to the Random Forest (RF) method. However, it doesn't even make the cut as one of the top ten most important features when assessed by other FS methods. On the other hand, EC(t) takes the spotlight as the most crucial feature within the mRMR and F-test algorithms. Interestingly, it secures the second and third most vital positions within the RF and Relief algorithms, respectively.

This discrepancy in feature importance highlights the critical role of ensemble FS techniques that leverage various filter and embedded FS methods. By embracing this diversity, we aim to harness the strengths of different approaches while mitigating the risk of selecting an unstable feature subset.

As a refinement to our research design, we exclusively incorporate the ten most crucial features identified through the ensemble FS process into the input feature subset for the ensemble filter-wrapper FS method. This adjustment aligns our methodology with established best practices in FS context.

The FS results for the ML models, which have been developed for the prediction of EC(t + 1) within the scope of this study, are presented in Table 6. In this table, the symbol "●" designates the inclusion of a specific feature, while "○" signifies the exclusion of the relevant feature.

After the identification of the ten most critical features (n) through the ensemble of mRMR, RF, F-test, and Relief FS methods, as depicted in Table 5 a final consensus is reached among them. The performance of each of the five ML models are evaluated using ten selected important features using the outcome feature subset (S1) obtained from ensemble filter method. The feature subset S1 includes EC(t), Coal, Gas, RnwS, Temp, Pop, P-Gas, SinM, CosM, and Mn#. Conversely, in the Wrapper method (S2), although a fewer number of features from all 18 features are typically selected on average, a diverse array of features is observed across different ML algorithms. It is noteworthy that, with the exception of P-Fuel, Coal, and Petr features, all other features find recommendation in at least one ML method. Interestingly, Mn#, CosM, P-El, and GDP features have been chosen by a minimum of four of the proposed ML algorithms, signifying their prominent importance and relevance.

In the hybrid filter-wrapper method (S1S2), a noticeable reduction is discerned in the number of selected features. On average, each of the proposed ML algorithms supports merely four features, except LSBoost which only three features are selected. Notably, Gas and CosM are two features that attain unanimous selection by all ML algorithms. Subsequently, SRnw and Mn# are the features that exhibit the most recurrent inclusion across the spectrum of ML algorithms.

The results depicted in Table 6 evoke a noteworthy trend, wherein the combination of both the filter and wrapper methods (S1S2) results in a more constrained selection of features. This suggests a higher level of consensus among the ML algorithms regarding the most relevant features, illustrated by the consistent selection of CosM and Gas. Furthermore, the recurrent inclusion of SRnw and Mn# underscores their enduring importance in the context of this predictive modeling exercise.

5.4. Final results

The results obtained from the application of the filter (S1), wrapper (S2), and filter-wrapper (S1S2) subsets for predicting EC(t + 1) are presented in Table 7, mirroring the information in Fig. 8a to 8c. To evaluate the ML algorithms more precisely, various assessment metrics, including RMSE, rRMSE, MAPE, MAE, and the coefficient of determination R2, have been utilized.

As shown in Table 6, the filter (S1) subset has identified 10 features as important, while the wrapper (S2) and filter-wrapper (S1S2) subsets have identified 7 and 4 features, respectively. Importantly, the results highlight the superior predictive accuracy of the wrapper (S2) subset for EC(t + 1). Among the ML algorithms within the wrapper subset (S2), the Support Vector Regression with Radial Basis Function (SVR-RBF)

Table 5
Results of ensemble FS scoring scheme for calculating average relative importance of 18 input features for target EC(t + 1).

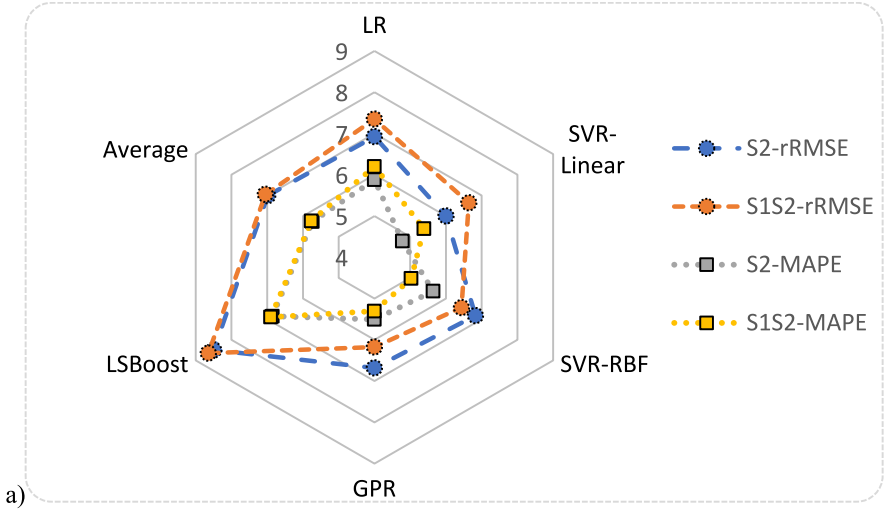
FS method	Real features															Month features			Total count
	EC(t)	Coal	Petr	Gas	B&W	El-Nuc	El-WSH	SRnw	GDP	Pop	Temp	EI	P-Gas	P-El	P-Fuel	SinM	CosM	Mn#	
mRMR	1	0.370	0.235	0.331	0.162	0	0.199	0.237	0	0	0.227	0	0.804	0	0	0	0.774	0	10
RF	0.740	0	0	0.452	0.519	0	0	0.670	0	0.551	0	0.705	0	0.616	0	0.705	1	0.787	10
F-test	1	0.630	0	0.800	0	0	0.240	0.256	0	0.242	0.406	0	0	0	0	0.319	0.813	0.885	10
Relief	0.412	0.147	0.447	0.529	0	0	0.088	0	0	0	0.201	0	0.386	0	0	0.259	1	0.358	10
Vote	4	3	2	4	2	0	3	3	0	2	3	1	2	1	0	3	4	3	40
Average importance	**0.788**	**0.287**	**0.170**	**0.528**	**0.170**	**0**	**0.132**	**0.291**	**0**	**0.198**	**0.208**	**0.176**	**0.297**	**0.154**	**0**	**0.320**	**0.897**	**0.507**	
Final rank	2	8	13	3	12	–	15	7	–	10	9	11	6	14	–	5	1	4	

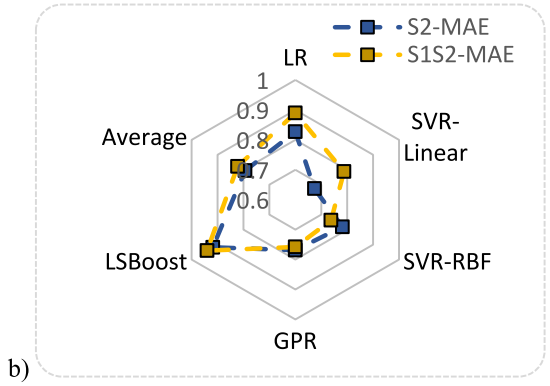
Table 6
Selected feature subsets from three FS approaches with five ML models for forecasting EC(t + 1).

Feature subset	ML models	Real features															Dummy features			Number of features
		EC (t)	Coal	Petr	Gas	B&W	El-Nuc	El-WSH	SRnw	GDP	Pop	Temp	EI	P-Gas	P-El	P-Fuel	SinM	CosM	Mn#	
Ensemble Filter (S1)	All five ML models	●	●	○	●	○	○	○	●	○	●	●	○	●	○	○	●	●	●	10
	Vote	5	5	0	5	0	0	0	5	0	5	5	0	5	0	0	5	5	5	10
Wrapper (S2)	LASSO	○	○	○	○	●	●	●	○	●	●	○	○	○	●	○	○	●	●	8
	SVR-Linear	○	○	○	○	●	○	○	○	●	●	○	○	○	●	○	○	●	●	6
	SVR-RBF	○	○	○	○	●	○	●	○	●	●	○	○	○	●	○	○	●	●	7
	GPR	●	○	○	●	○	○	○	○	●	○	●	○	●	●	○	●	○	●	8
	LSBoost	●	○	○	○	○	○	○	●	○	○	●	●	○	○	○	○	●	●	6
	Vote	2	0	0	1	3	1	2	1	4	3	2	1	1	4	0	1	4	5	7
Ensemble Filter-Wrapper (S1S2)	LASSO	○	○		●				●		○	○		○			○	●	●	4
	SVR-Linear	○	○		●				●		○	○		○			○	●	●	4
	SVR-RBF	○	○	*N/A*	●	*N/A*	*N/A*	*N/A*	●	*N/A*	○	○	*N/A*	○	*N/A*	*N/A*	○	●	●	4
	GPR	○	○		●				○		○	●		○			●	●	○	4
	LSBoost	○	○		●				●		○	○		○			○	●	○	3
	Vote	0	0		5				4		0	1		0			1	5	3	3.8

Table 7
Performance metrics of ML models for forecasting EC(t + 1) with ensemble filter (S1), wrapper (S2) and ensemble filter-wrapper (S1S2) FS approaches.

Feature subset	ML models	Metrics					Nubmer of features
		RMSE	rRMSE	MAPE	MAE	R2	
	LASSO	1.5774	10.4614	8.2645	1.2061	0.8136	10
	SVR-Linear	1.6398	10.8755	8.5298	1.2599	**0.8219**	10
Ensemble filter (S1)	SVR-RBF	1.6048	10.6434	8.188	1.1031	0.7204	10
	GPR	**1.5540**	**10.3060**	**7.9540**	**1.0683**	0.7201	10
	LSBoost	1.6237	10.7683	8.4587	1.1403	0.7511	10
Average		1.4896	9.8788	7.7809	1.0775	0.7725	10
	LASSO	1.0439	6.9234	5.878	0.827	0.8436	8
	SVR-Linear	1.0328	6.8493	5.7961	0.8099	0.8481	**6**
Wrapper (S2)	SVR-RBF	**0.9051***	**6.0024***	**4.7836***	**0.6740***	**0.8870***	7
	GPR	1.0075	6.6820	5.5061	0.7686	0.8620	8
	LSBoost	1.2810	8.4956	6.8566	0.9182	0.8354	**6**
Average		1.0534	6.9860	5.7336	0.7939	0.8554	7
	LASSO	1.1079	7.348	6.2008	0.8896	**0.8864**	4
	SVR-Linear	1.0014	6.6411	5.3864	0.7878	0.8810	4
Ensemble filter-Wrapper (S1S2)	SVR-RBF	0.9706	6.4370	**5.0280**	**0.7376**	0.8664	4
	GPR	**0.9312**	**6.1755**	5.3049	0.7579	0.8776	4
	LSBoost	1.3038	8.6468	6.9009	0.9402	0.8130	**3***
Average		1.0630	7.0497	5.7642	0.8226	0.8649	3.8





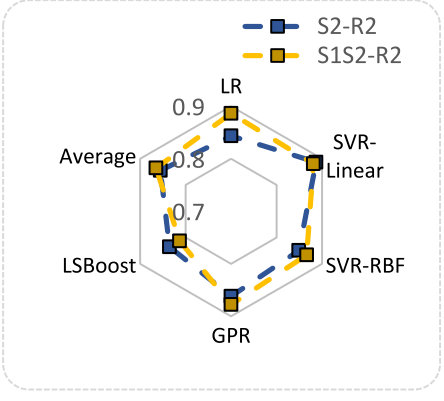


Fig. 8. Performance metrics of the ML models for S2 and S1S2 feature subset for forecasting EC(t + 1) a) rRMSE and MAPE, b) MAE, and c) R2.

algorithm has proven to be the most proficient, demonstrating superior accuracy compared to other algorithms. It has achieved notable performance metrics with values of 0.9051, 6.0024, 4.7836, 0.6740, and 0.8870 in RMSE, rRMSE, MAPE, MAE, and R2, respectively.

Figs. 9 and 10 present a detailed representation of the training and testing processes of the SVR-RBF model used for predicting EC(t + 1). These processes occur in the context of feature subsets referred to as S1S2 and S2.

The SVR-RBF model, coupled with both the wrapper and ensemble filter-wrapper feature selection (FS) approaches, consistently exhibits superior predictive performance. This is evident in the attainment of low error metrics and a substantial degree of explained variance. Collectively, these metrics underscore the model's ability to deliver highly accurate forecasts of EC(t + 1).

A noteworthy observation from these findings pertains to the minimal decline in performance when utilizing a smaller feature subset.

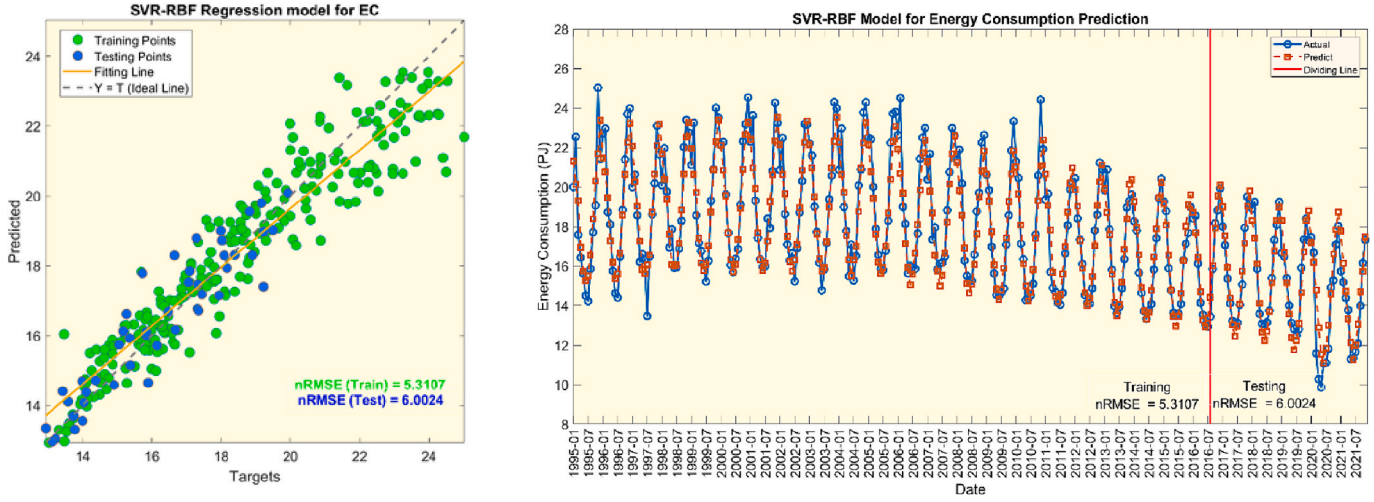


Fig. 9. Performance of SVR-RBF in forecasting EC(t + 1) based on wrapper FS (S2) (a) regression model, b) time series plot.

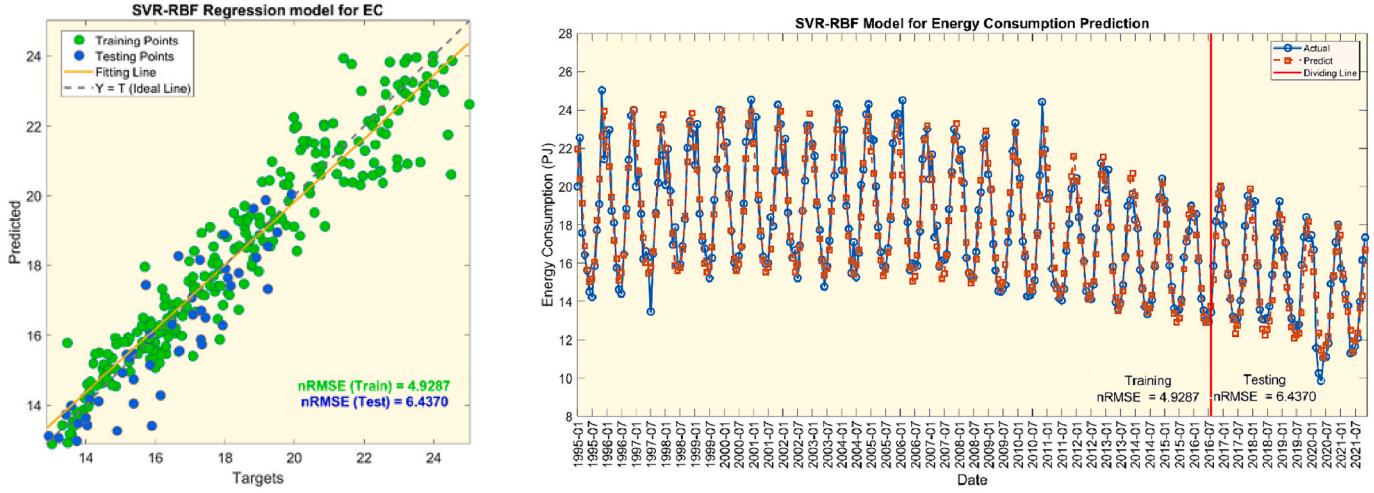


Fig. 10. Performance of SVR-RBF in forecasting EC(t + 1) based on filter-wrapper FS (S1S2).

Specifically, S2 encompasses three additional features compared to S1S2 in predicting EC(t + 1). This outcome emphasizes the effectiveness of the hybrid ensemble filter-wrapper FS approach proposed in this study, successfully reducing dimensionality while preserving essential evaluation metrics within the forecasting process.

Additionally, a careful assessment of SVR's performance on the test set reveals a slight decrease in its performance compared to the training set. This observation suggests a minor tendency toward overfitting, indicating that the model may have a slight inclination to fit the training data too closely. It's important to emphasize, however, that the extent of this overfitting is limited, emphasizing the model's strength and its ability to generalize effectively.

5.5. SHAP analysis

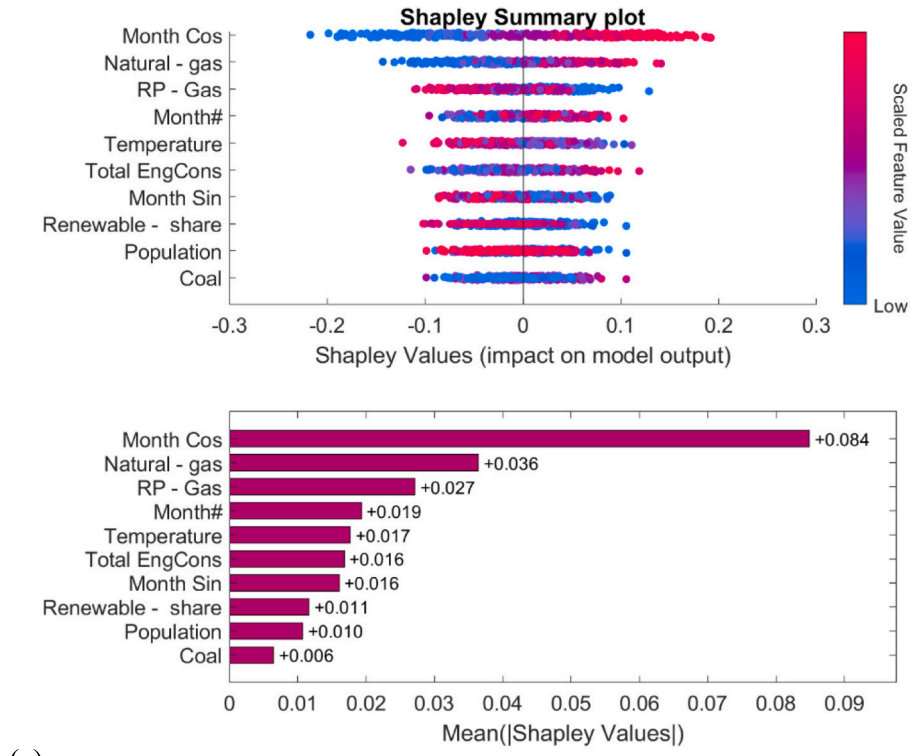
Figs. 11 to 13 detail SHAP values for feature subsets from top ML models (SVR-RBF and GPR) forecasting EC(t + 1). In this study, all features are continuous and vertically sorted by average impact on EC(t + 1) predictions. Month Cos holds the highest impact with varying influence in both ML models across all three FS approaches, except GPR with wrapper (S2). The primary reason for such a significant impact could be the corresponding fluctuations of the Cosine function of Month with the target feature EC(t + 1). This suggests that the most effective explanatory feature for forecasting EC(t + 1) is simply the Month, which is effectively represented by the Cosine of the Month.

The remaining nine features for SVR-RBF and GPR with ensemble filter (S1) exhibit notable consistency, with one exception: Natural gas holds the second most impactful position for SVR-RBF but is ranked tenth for GPR.

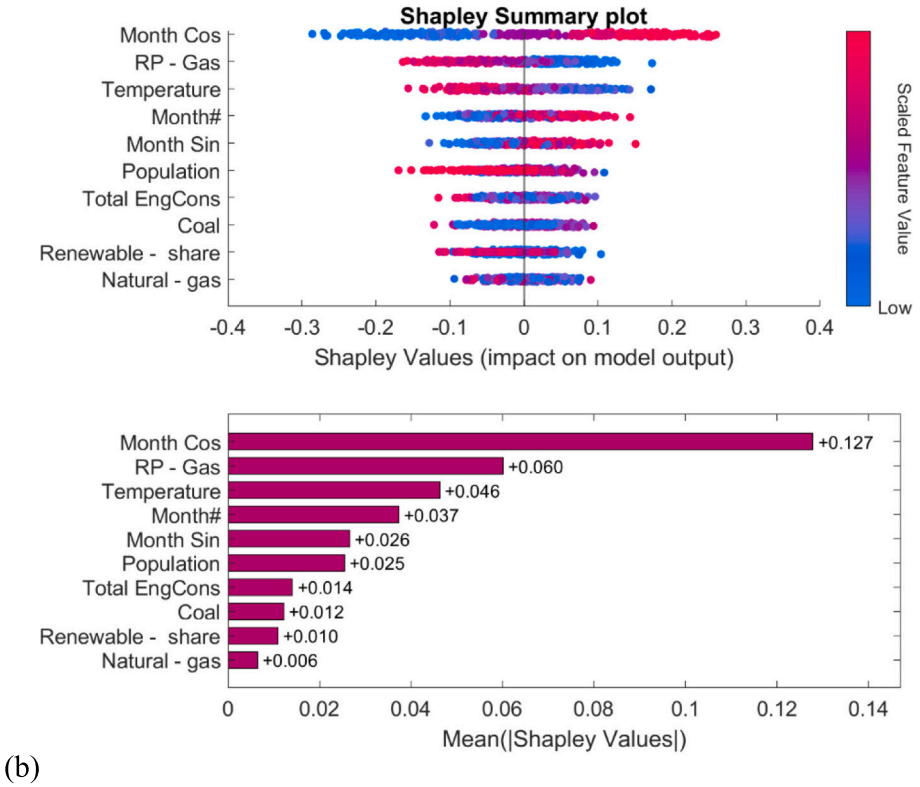
In Fig. 12(a), SVR-RBF with wrapper, the top-performing model among all 15 configurations, highlights RP-Electricity, Temperature, and Bioenergy & waste as the second, third, and fourth most influential features, respectively. In Fig. 12(b), GPR with wrapper showcases Month#, RP-Gas, Month Sin, and Population as the first, second, third, and fourth most influential features, respectively.

Ranked as the 2nd and 3rd best-performing models, GPR and SVR-RBF with ensemble filter-wrapper strike a balance by offering optimal compromise metrics across various evaluation criteria and a minimal number of features. Fig. 13 reveals both commonalities and distinctions between these models. Each employs only four features to generate forecasts. For SVR-RBF, the second and third most impactful features are Share of renewable and Natural gas, while for GPR, it is Population and Natural gas, respectively.

The majority of feature contributions align with intuitive expectations and UK government policies, as illustrated in Fig. 6. In other words, high values of Share of renewable, Bioenergy & waste, Temperature RP-Electricity, RP-Gas have negative impact on EC(t + 1) forecasts, while high values of Natural gas, Month Cos and Month# have positive impact



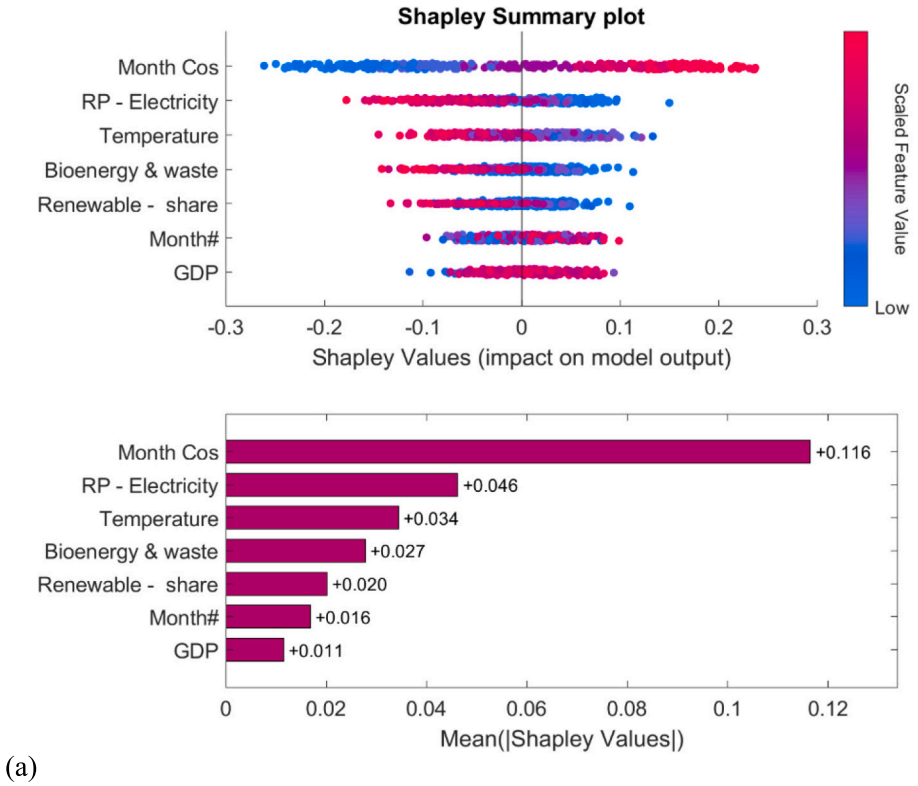
(a)



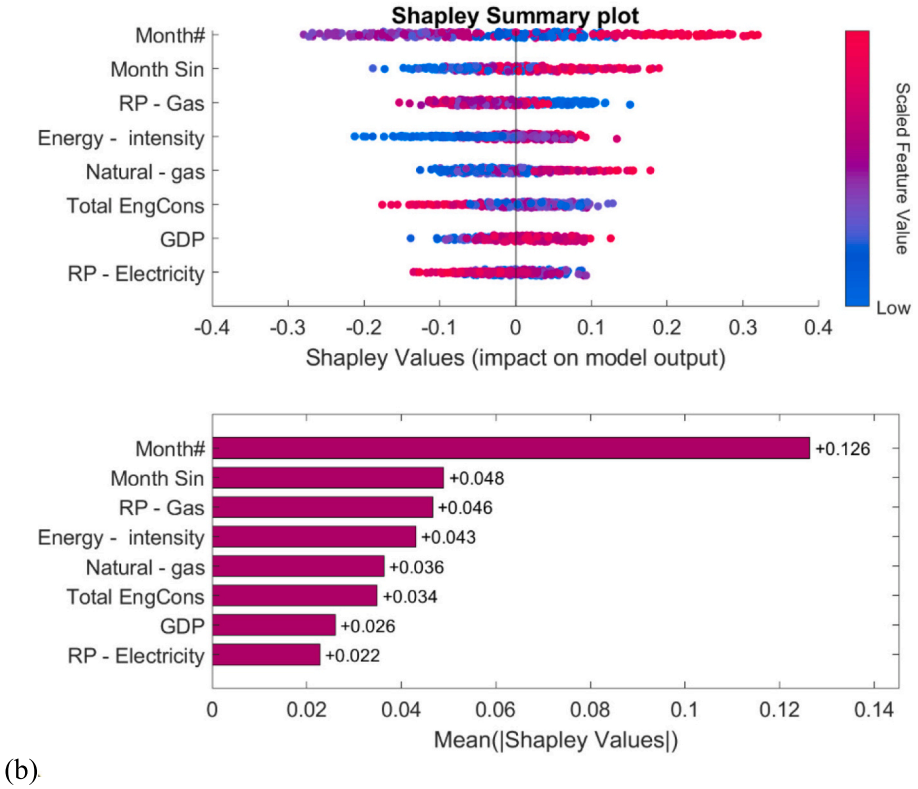
(b)

Fig. 11. The Shapley analysis including Shapely summary plot and the average contributions of the ensemble filter feature subset (S1) for forecasting EC(t + 1): (a) SVR-RBF, (b) GPR.

on EC(t + 1) forecasts. However, three features—RP-Gas, Temperature, and Share of renewable—play a distinct role in influencing the forecasts of these ML models during global interpretation. This unique impact may stem from multicollinearity and the substantial influence of other features on EC(t + 1), a common occurrence in various ML applications.



(a)

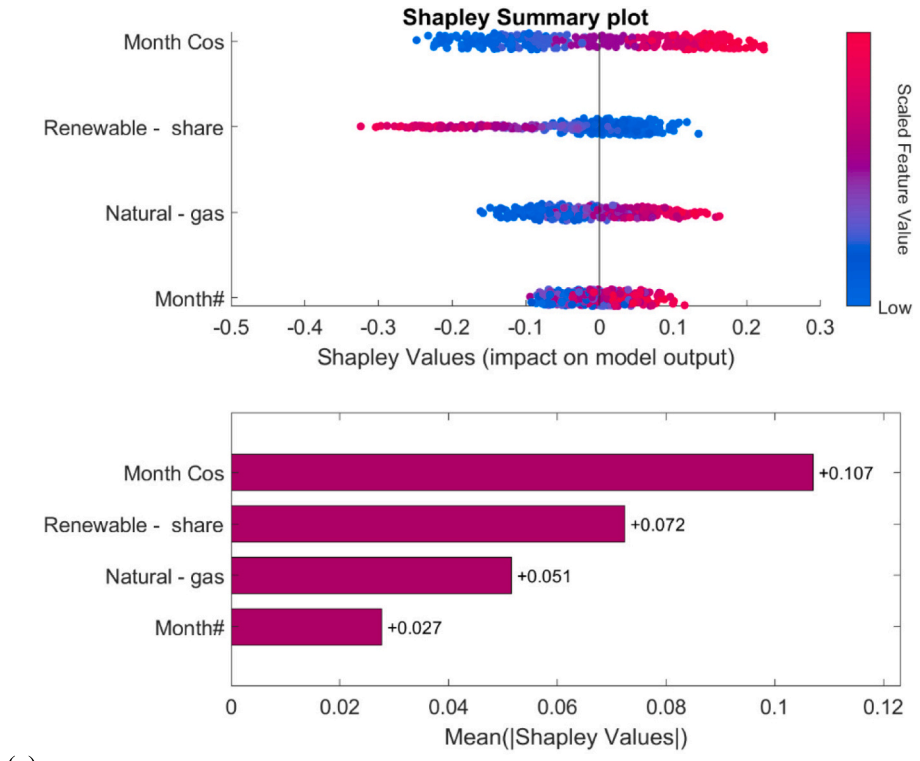


(b)

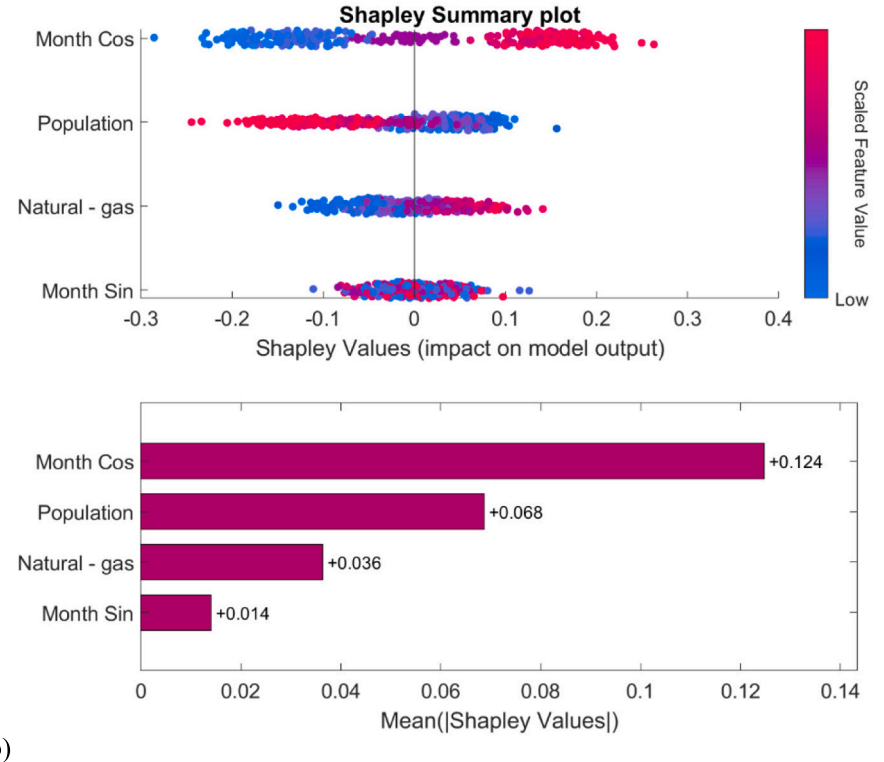
Fig. 12. The Shapley analysis including Shapely summary plot and the average contributions of the wrapper feature subset (S2) for forecasting EC(t + 1): (a) SVR-RBF, (b) GPR.

6. Conclusions and future research

This research introduces a pioneering framework that seamlessly integrates feature selection (FS) and machine learning (ML) techniques, representing a significant advancement in the realm of predicting inland national energy consumption (EC) in the United Kingdom. The framework strategically combines three distinctive FS approaches—ensemble filter, wrapper, and ensemble filter-wrapper—with five interpretable ML



(a)



(b)

Fig. 13. The Shapley analysis including Shapely summary plot and the average contributions of the ensemble filter-wrapper feature subset (S1S2) for forecasting EC (t + 1): (a) SVR-RBF, (b) GPR.

models—LASSO, SVR-Linear, SVR-RBF, GPR, and LSBoost. This integration not only allows us to identify the ML models with the highest forecast accuracy but also guarantees transparency in the forecasting process. The framework's strength lies in its ability to calculate the relative importance of contributing features and their impact on the target EC(t + 1). This dual focus on model accuracy and interpretability is a distinctive feature that enhances the reliability and utility of the forecasting process.

This study marks a multifaceted contribution to ML-based EC forecasting. Notably, it advances the use of high-frequency monthly EC data

at the national level, providing fresh insights into multi-source EC statistics in the UK. This departure from the reliance on annual data, as observed in other studies, strengthens the reliability and robustness of our findings. The comprehensive consideration of various features, including energy sources, economic and demographic factors, climate variables, and resource and production factors, sets this study apart from prior research. The introduction of a novel generalizable ensemble FS approach, the meticulous and comparative FS analysis, and the incorporation of Shapley additive explanation (SHAP) analysis contribute to the uniqueness and significance of this research. The proposed framework not only improves prediction accuracy but also enhances the interpretability of black-box ML models, making it a valuable tool for policymakers and energy analysts. The framework's interpretability is underscored by its ability to elucidate the relative importance of features, facilitating a more nuanced understanding of their contributions to EC forecasting.

The experimental findings reveal that, among the 15 configurations tested, the combined SVR-RBF model with the wrapper FS approach consistently exhibits the most robust predictive performance across all five evaluation metrics. Notably, in the context of both ensemble filter and ensemble filter-wrapper FS approaches, GPR consistently outperforms SVR-RBF for the majority of metrics. An intriguing observation emerges regarding the minimal decline in the performance of ML models when utilizing a smaller feature subset. Specifically, the feature subset derived from the wrapper approach (S2) includes, on average, three additional features compared to the feature subset from the ensemble filter-wrapper approach (S1S2) in predicting EC(t + 1). This outcome underscores the efficacy of the hybrid ensemble filter-wrapper FS approach proposed in this study by successfully reducing dimensionality while preserving essential performance in the forecasting process.

In essence, this study's innovative framework stands as a noteworthy contribution to the field of EC forecasting. By seamlessly integrating three FS approaches and interpretable ML techniques, it addresses both accuracy and interpretability concerns, making it a valuable tool for advancing the understanding and prediction of inland national energy consumption in the United Kingdom.

Looking forward, additional research is imperative to enhance and address limitations in the proposed framework. A critical aspect requiring attention is the development of a more effective FS approach to discern and select the most influential features. Furthermore, there is a need to explore alternative interpretable ML methods, such as Gradient-weighted Class Activation Mapping (Grad-CAM), within the context of time series forecasting. Subsequent investigations could extend the current work by comparing the outcomes with diverse FS approaches and interpretable ML models, thereby evaluating the performance achieved in this study. This comparative analysis would offer a comprehensive understanding of the effectiveness of various FS approaches and contribute nuanced insights into the role and potential of interpretable ML methods within the time series forecasting domain.

CRediT authorship contribution statement

Hamidreza Eskandari: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hassan Saadatmand:** Visualization, Validation, Software, Methodology. **Muhammad Ramzan:** Writing – original draft. **Mobina Mousapour:** Writing – original draft.

Declaration of competing 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.

Data availability

Data will be made available on request.

Appendix A. Appendix

Table A1
Descriptive statistics for the fifteen main features.

Feature	Unit	Mean	Std	Min	Max	Source
Total monthly final EC	Mtoe	17.76	3.27	9.86	25.02	UK-ESNZ
Coal	Mtoe	2.64	1.27	0.34	6.19	UK-ESNZ
Petroleum	Mtoe	5.91	0.65	2.57	7.63	UK-ESNZ
Natural gas	Mtoe	6.91	2.24	3.11	11.83	UK-ESNZ
Bioenergy & waste	Mtoe	0.63	0.46	0.14	1.69	UK-ESNZ
Electricity - nuclear	Mtoe	1.40	0.36	0.70	2.46	UK-ESNZ
Electricity – wind, solar & hydro	Mtoe	0.20	0.22	0.01	0.94	UK-ESNZ
Share of renewable energy	%	0.05	0.05	0.01	0.19	MaCal
GDP per capita	USD/person	38,186.13	7852.35	21,975.20	50,918.10	UK-ONS
Population	Million	62.30	3.09	58.02	67.08	UK-ONS
Average Temperature	Celsius	10.35	4.46	−0.27	19.26	UK-ESNZ
Energy intensity per unit of GDP (SDG 7.3)	TES/GDP	3.74	0.90	2.25	5.12	IEA
Real price - Gas	USD	85.48	27.46	50.80	131.70	UK-ESNZ
Real price- Electricity	USD	97.15	23.70	65.00	148.30	UK-ESNZ
Real price - Motor fuel & oil	USD	86.89	13.28	61.00	117.60	UK-ESNZ

Note: UK-ESNZ: UK Department for Energy Security and Net Zero, UK-ONS: UK Office for National Statistics, and IEA: International Energy Agency, MaCal = Manually calculated from fuels consumption data (first seven features) provided by UK-ESNZ.

References

[1] Pao H-T, Fu H-C, Tseng C-L. Forecasting of CO2 emissions, energy consumption and economic growth in China using an improved grey model. Energy 2012;40(1):400–9.
[2] Maaouane M, Zouggar S, Krajačić G, Zahboune H. Modelling industry energy demand using multiple linear regression analysis based on consumed quantity of goods. Energy 2021;225:120270.
[3] Liu Y, Li J. Annual electricity and energy consumption forecasting for the UK based on back propagation neural network, multiple linear regression, and least square support vector machine. Processes 2022;11(1):44.
[4] Sharifzadeh M, Sikinioti-Lock A, Shah N. Machine-learning methods for integrated renewable power generation: a comparative study of artificial neural networks, support vector regression, and Gaussian process regression. Renew Sust Energ Rev 2019;108:513–38.

[5] Ağbulut Ü. Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. Sustain Product Consumpt 2022;29:141–57. https://doi.org/10.1016/j.spc.2021.10.001.
[6] Qin Y, Ke J, Wang B, Filaretov GF. Energy optimization for regional buildings based on distributed reinforcement learning. Sustain Cities Soc 2022;78:103625.
[7] Aras S, Hanifi Van M. An interpretable forecasting framework for energy consumption and CO2 emissions. Appl Energy 2022;328:120163. https://doi.org/10.1016/j.apenergy.2022.120163.
[8] Kamani D, Ardehali M. Long-term forecast of electrical energy consumption with considerations for solar and wind energy sources. Energy 2023;268:126617.
[9] Li Z, Zhou B, Hensher DA. Forecasting automobile gasoline demand in Australia using machine learning-based regression. Energy 2022;239:122312.
[10] Chen Z, Xiao F, Guo F, Yan J. Interpretable machine learning for building energy management: A state-of-the-art review. Adv Appl Energy 2023:100123.
[11] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st international conference on neural information processing systems, Long Beach, California, USA. 2017.
[12] Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Front Bioinform 2022;2:927312. https://doi.org/10.3389/fbinf.2022.927312.
[13] Theng D, Bhoyar KK. Feature selection techniques for machine learning: a survey of more than two decades of research. Knowl Inf Syst 2024;66:1575–637. https://doi.org/10.1007/s10115-023-02010-5.
[14] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3(Mar):1157–82.
[15] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing 2018;300:70–9.
[16] Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J Appl Sci Technol Trends 2020;1(2):56–70.
[17] Zyl C, Ye X, Naidoo R. Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: a comparative analysis of Grad-CAM and SHAP. Appl Energy 2024;353:122079. https://doi.org/10.1016/j.apenergy.2023.122079.
[18] Qiao Q, Eskandari H, Saadatmand H, Sahraei MA. An interpretable multi-stage forecasting framework for energy consumption and CO2 emissions for the transportation sector. Energy 2024;129499.
[19] Hoxha J, Çodur MY, Mustafaraj E, Kanj H, El Masri A. Prediction of transportation energy demand in Türkiye using stacking ensemble models: methodology and comparative analysis. Appl Energy 2023;350:121765.
[20] Lv S-X, Wang L. Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model. Energy 2023;263:126100.
[21] Wang C, Cao Y. Forecasting Chinese economic growth, energy consumption, and urbanization using two novel grey multivariable forecasting models. J Clean Prod 2021;299:126863.
[22] Javanmard ME, Tang Y, Wang Z, Tontiwachwuthikul P. Forecast energy demand, CO2 emissions and energy resource impacts for the transportation sector. Appl Energy 2023;338:120830.
[23] Nazir A, Shaikh AK, Shah AS, Khalil A. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (TFT). Results Eng 2023;17:100888.
[24] Ahmed AAM, Bailek N, Abualigah L, Bouchouicha K, Kuriqi A, Sharifi A, et al. Global control of electrical supply: a variational mode decomposition-aided deep learning model for energy consumption prediction. Energy Rep 2023;10:2152–65.
[25] Rao C, Zhang Y, Wen J, Xiao X, Goh M. Energy demand forecasting in China: a support vector regression-compositional data second exponential smoothing model. Energy 2023;263:125955.
[26] Saglam M, Spataru C, Karaman OA. Forecasting electricity demand in Turkey using optimization and machine learning algorithms. Energies 2023;16(11):4499.
[27] Atems B, Mette J, Lin G, Madraki G. Estimating and forecasting the impact of nonrenewable energy prices on US renewable energy consumption. Energy Policy 2023;173:113374. https://doi.org/10.1016/j.enpol.2022.113374.
[28] Khan AM, Osińska M. Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India. Expert Syst Appl 2023;212:118840.
[29] Aslan M, Beşkirli M. Realization of Turkey's energy demand forecast with the improved arithmetic optimization algorithm. Energy Rep 2022;8:18–32. https://doi.org/10.1016/j.egyr.2022.06.101.
[30] Maaouane M, Chennaif M, Zouggar S, Krajačić G, Duić N, Zahboune H, et al. Using neural network modelling for estimation and forecasting of transport sector energy demand in developing countries. Energy Convers Manag 2022;258:115556. https://doi.org/10.1016/j.enconman.2022.115556.
[31] Sahraei MA, Çodur MK. Prediction of transportation energy demand by novel hybrid meta-heuristic ANN. Energy 2022;249:123735. https://doi.org/10.1016/j.energy.2022.123735.
[32] Zheng C, Wu W-Z, Xie W, Li Q, Zhang T. Forecasting the hydroelectricity consumption of China by using a novel unbiased nonlinear grey Bernoulli model. J Clean Prod 2020;123903. https://doi.org/10.1016/j.jclepro.2020.123903.
[33] Ahmed YA, Koçer B, Huda S, Saleh Al-rimy BA, Hassan MM. A system call refinement-based enhanced minimum redundancy maximum relevance method for ransomware early detection. J Netw Comput Appl 2020;167:102753. https://doi.org/10.1016/j.jnca.2020.102753.
[34] Wen Y, Wu R, Zhou Z, Zhang S, Yang S, Wallington TJ, et al. A data-driven method of traffic emissions mapping with land use random forest models. Appl Energy 2022;305:117916. https://doi.org/10.1016/j.apenergy.2021.117916.
[35] Kira K, Rendell LA. A practical approach to feature selection. In: Sleeman D, Edwards P, editors. Machine learning proceedings 1992. Morgan Kaufmann; 1992. p. 249–56. https://doi.org/10.1016/B978-1-55860-247-2.50037-1.
[36] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn 2003;53(1):23–69. https://doi.org/10.1023/A:1025667309714.
[37] Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Comput Biol Med 2019;112:103375. https://doi.org/10.1016/j.compbiomed.2019.103375.
[38] Bolón-Canedo V, Alonso-Betanzos A. Ensembles for feature selection: a review and future trends. Inf Fusion 2019;52:1–12. https://doi.org/10.1016/j.inffus.2018.11.008.
[39] Wang Q, Chen D, Li M, Li S, Wang F, Yang Z, et al. A novel method for petroleum and natural gas resource potential evaluation and prediction by support vector machines (SVM). Appl Energy 2023;351:121836. https://doi.org/10.1016/j.apenergy.2023.121836.
[40] Wang J, Xu P, Xiaobo Ji M, Li W Lu. MIC-SHAP: an ensemble feature selection method for materials machine learning. Mater Today Commun 2023;37:106910. https://doi.org/10.1016/j.mtcomm.2023.106910.
[41] Almalaq A, Zhang JJ. Evolutionary deep learning-based energy consumption prediction for buildings. IEEE Access 2019;7:1520–31. https://doi.org/10.1109/ACCESS.2018.2887023.
[42] Zhou Q, Li Y, Zhao D, Li J, Williams H, Xu H, et al. Transferable representation modelling for real-time energy management of the plug-in hybrid vehicle based on k-fold fuzzy learning and Gaussian process regression. Appl Energy 2022;305:117853. https://doi.org/10.1016/j.apenergy.2021.117853.
[43] Centofanti F, Fontana M, Lepore A, Vantini S. Smooth LASSO estimator for the function-on-function linear regression model. Comput Stat Data Anal 2022;176:107556. https://doi.org/10.1016/j.csda.2022.107556.
[44] Huang J-C, Tsai Y-C, Wu P-Y, Lien Y-H, Chien C-Y, Kuo C-F, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput Methods Prog Biomed 2020;195:105536. https://doi.org/10.1016/j.cmpb.2020.105536.
[45] Guven D, Kayalica MO. Analysing the determinants of the Turkish household electricity consumption using gradient boosting regression tree. Energy Sustain Dev 2023;77:101312. https://doi.org/10.1016/j.esd.2023.101312.
[46] Awal Md A, Masud M, Hossain Md S, Bulbul AA-M, Mahmud SMH, Bairagi AK. A novel Bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. IEEE Access 2021;9:10263–81.
[47] Turner R, Eriksson D, McCourt M, Kiili J, Laaksonen E, Xu Z, et al. Bayesian optimization is superior to random search for machine learning hyperparameter tuning: analysis of the black-box optimization challenge 2020. In: Proceedings of the NeurIPS 2020 competition and demonstration track, virtual event/Vancouver, BC, Canada, 6–12 December. vol. 133; 2020. p. 3–26.
[48] Saadatmand H, Akbarzadeh-T M-R. Set-based integer-coded fuzzy granular evolutionary algorithms for high-dimensional feature selection. Appl Soft Comput 2023;142:110240.
[49] Ye M, Li L, Yoo D-Y, Li H, Zhou C, Shao X. Prediction of shear strength in UHPC beams using machine learning-based models and SHAP interpretation. Constr Build Mater 2023;408:133752. https://doi.org/10.1016/j.conbuildmat.2023.133752.
[50] Kaya M, Baha Karan M, Telatar E. Electricity price estimation using deep learning approaches: an empirical study on Turkish markets in normal and Covid-19 periods. Expert Syst Appl 2023;224:120026. https://doi.org/10.1016/j.eswa.2023.120026.
[51] Qiao Q, Yunusa-Kaltungo A, Edwards RE. Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and empirical mode decomposition. Energy Rep 2023;9:3643–60. https://doi.org/10.1016/j.egyr.2023.02.046.
[52] EViews.. EViews 12 User's guide. IHS Markit; 2022.