ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Energy Consumption Prediction Using Extreme Random Forest and Hierarchical Sampling

Jianqi Yin[1], *Shilai Yuan*[1] *and Minjie Zhu*[1]

[1] Hangzhou Cigarette Factory, China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou, 310024, Zhejiang, China

SUMMARY: *To solve the prediction accuracy is not enough because the energy consumption sequence of the tobacco industry has a high degree of volatility and multi-modality. This study proposes an integrated prediction model based on improved hierarchical sampling and extreme random forest to overcome the inherent high volatility and multi-modal characteristics of energy consumption sequences in the tobacco industry that cause low accuracy in predicting models. During data processing, we developed a more advanced hierarchical sampling method which included multi-dimensional combination stratification and key event weighting, improving sample quality. Model building, traditional random forests go through dual progressive optimization, first adding weighted feature selection then building up a segmental point randomizing mechanism for making it into an extremely randomized tree. In the validation of the sampling method, improved hierarchical sampling covered 95.5% of important events with only 218 samples, but just 56.1% were covered by simple random sampling. Energy consumption forecast using the extreme random forest model predicted about 5850 kW·h around day 10 during the energy consumption trough, which was close to the real 5900 kW·h. Production Day - Heating Season operating condition. The AAE of the Final Energy Consumption Prediction Model is 6.1%. The proposed model captures energy consumption dynamics in complicated tobacco plant operations, providing technical support for companies transitioning from empirical scheduling to data-driven precise energy management decisions.*

KEYWORDS: *virtual reality; English immersive teaching; scenario construction; intercultural communicative competence; higher education*

1 Introduction

In the process of promoting energy efficiency, productivity enhancement and intelligent upgrading in the context of Industry 4.0 and energy structure optimization, this becomes a core part of the country's energy strategy[1-3]. In terms of traditional manufacturing enterprises such as tobacco companies, which are important consumers of energy, it needs a lot of stable sources like steam and electricity during the process of producing things, so that costs related to power usage can be quite large compared with other types [4]. But usually, their company has very complicated kind of systems for giving out energy, where changing amounts of energy they use depends on what kind of plan they have at different times when making products along with turning machines off/on together. On one hand these kinds of adjustments happen all over again throughout production plans that create frequent shifts between how much we need energy right now versus before;[5] And also because there are some outside forces too - like weather or

*17300983536@163.com
https://doi.org/10.65102/is2026784
Published: 30 April, 2026

holidays - that add more unpredictability to how much load we might see coming up later on![6] It is hard to balance supply and demand because there are many different ways to produce something. Too much supply means that you are using up energy, too little will interfere with your production plan and impact profits. So we need smart models to accurately predict future electricity consumption and manage the company's electricity well.

It has become a popular topic in Industrial Engineering and Data Science to accurately predict and optimize energy consumption under different complicated situations in recent years, Machine Learning is one representative data driven approach that can develop. Like Zuo Hao et al., who made a systematic comparison of various machine learning models used for industrial energy consumption prediction and pointed out that the ensemble method of random forest was very beneficial for solving computational complexity and accuracy problems of the traditional model [7]. Scholars like Ullah compare advanced models such as Extreme Gradient Boosting (XGBoost) and light gradient boosting methods which perform better than the traditional linear regression and neural networks on prediction accuracy and robustness for complicated energy use [8]. However, when these general-purpose models are applied directly to the tobacco manufacturing industry, they often encounter some constraints because of the unique "highly volatile and multimodal" energy consumption characteristics of tobacco production. While current research mainly focuses on certain types of industries, their operating environments differ greatly from those in tobacco factories. For example, Yi Wangyuan's team proposed a DBO-BP framework improved by Dung Beetle Optimization Algorithm to tackle the issues related to CNC milling machines' energy consumption forecasting and parameter optimization, thereby achieving energy savings; however, since the CNC milling machine's operation is fairly monotonous, this model can't be easily adapted into capturing the intricacies within the production process of making cigarettes. [9] Summary: Current research on industrial energy consumption prediction has achieved some progress, but most methods cannot capture the nonlinear data characteristics in complex scenarios such as tobacco factories, or have shortcomings in sample representativeness and model generalization ability. In order to solve these problems, we will propose a forecasting model that integrates improved hierarchical sampling and Extremely Randomized Trees (ERT) in order to achieve an accurate and widely applicable forecasting model that can accurately predict the amount of energy consumed during complex operations at tobacco facilities.

Study is mainly for the air conditioning system of cigarette production, it takes a large proportion of energy consumption and has great influence from the combination of production and climate. The energy consumption accounts for about 35% to 50% of all energy used by plants. It provides a stable environment with uniform temperatures and humidity levels throughout core workshops like cigarette rolling, packaging, and tobacco processing. Its complex operation logic is also one of the reasons why the entire factory's energy usage pattern is superimposed on multiple modes and non-linearly.

Innovation of this study is not simply an algorithm applied but rather continuous improvement across all levels - data processing & modeling. In terms of data, we make the standard hierarchical sampling into a more refined sampling strategy that has multidimensional combinations. and weighted key events at the data level. At the model level, it improves on the conventional random forest algorithm with two mechanisms–weighting feature selection and randomized segmentation points–to produce an enhanced version called the extreme random forest. The establishment of such integrated forecasting models can give technical support for precise energy management and low-cost energy saving for tobacco, similar manufacturers and other industries.

2 Construction of an Energy Consumption Prediction Model

2.1 Progressive Improvement of Data Processing for Hierarchical Sampling

Tobacco company energy consumption has many overlapping modes and nonlinear fluctuations, making it difficult to achieve reliable energy consumption forecasts [10, 11]. On the one hand, energy consumption is quite different on a day when it's being made compared to days it's not. The loads of the seasons (heating in winter and cooling in summer) are also changing all the time together with what's happening at the production site, so they form something called a "composite pattern". So developing an advanced data processing method that can fully make use of the intrinsic data structure and improve the model's generalization ability is very important. Traditional ML approaches typically use basic random sampling for splitting the dataset; however, such an approach neglects the data's heterogeneity across different operating conditions [12]. In tobacco factory and so on such a place, energy consumption changes greatly during day when it works or not and heating or no heating. Simple random sampling will cause samples from these sets to be imbalanced, resulting in poor prediction. Hence, this paper introduces a hierarchical sampling technique as its main method of processing data. It splits the dataset into several similar "layers", each one is sampled separately while keeping the balance of training examples. Specifically, the quantity of samples taken out of every level following some kind of allocation plan will appear like Equation 1[14].

$$n_h = n \cdot \frac{N_h}{N} \quad (h = 1, 2, \ldots, L) \tag{1}$$

In Equation , represents the number of samples to be drawn from the N_h ith layer N_h; h represents the total N_h planned sample size N; represents Lthe total sample size L of the ith layer in the dataset;Represents all the samples in the entire dataset; and represents the number of layers. Equation (1) guarantees macro-sample balance. However, for tobacco factory scenario that energy consumption pattern is decided by multiple factor coupling, only single-dimensional stratification granularity will be too course to capture those complex synergies. The study improves on standard stratified sampling using multi-dimensional composite stratification. The "production planning" and "seasonal climate" as two important factors were decomposed into several orthogonal "composite layers", shown in Figure 1.

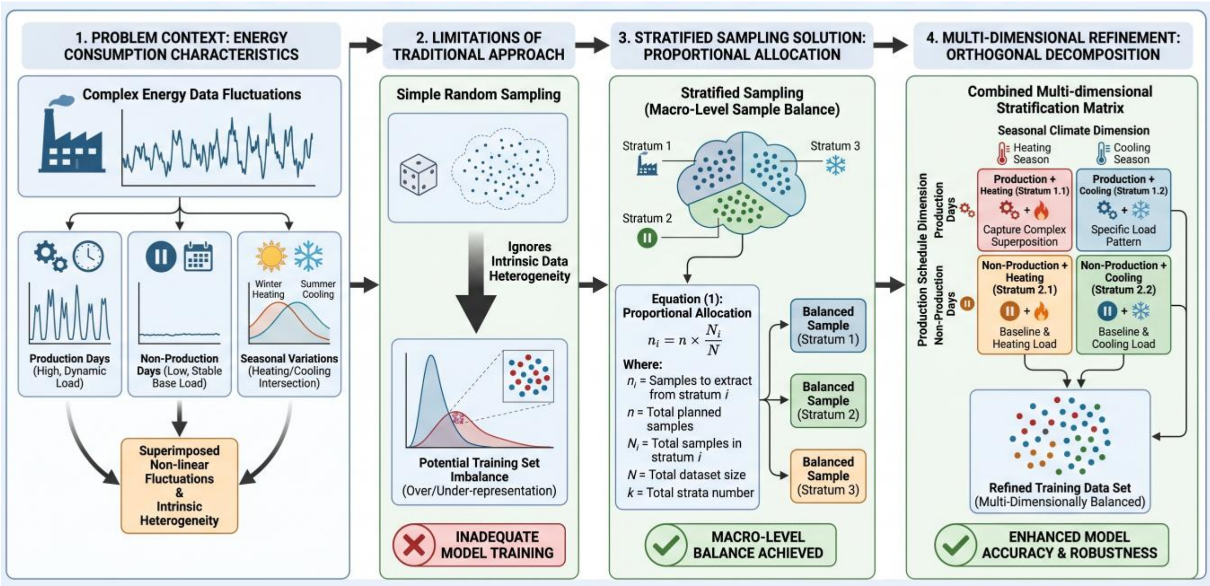


Figure 1: Multidimensional combined hierarchical schematic diagram based on tobacco factory operational modes

Figure 1 shows the specific steps for multi-dimensional combinatorial stratification. First, consolidate the original energy consumption dataset that integrates annual production records and meteorological consumption information. Afterward, based on the macro-level "seasonal climate" dimension, we initially split all annual data into two parts: "heating season" and "non-heating season." On this basis, each set is again subdivided by "production dimension", forming four more refined and internally consistent combination levels: "Production day - Heating season," Non-production day - Heating season," Production day - Non-heating season," Non-production day - Non-heating season". Finally, according to the proportionality of these four combinatorial layers, training sets and test sets are established to fully demonstrate the complicated operation mode of tobacco factory. Original dataset is divided into 4 non-overlapping combinatorial layer, as per Eq (2):

$$D = D_{ps} \cup D_{pn} \cup D_{nps} \cup D_{npn} \quad (2)$$

In Equation (D_{ps} 2), represents the "production D_{pm} day–heating season D_{pns} " layer, represents D_{pns} the "production D_{npn} day– D_{npn} non-heating season" layer, It means "non-production day-heating season" layer, It means "non-production day-non-heating season" layer. Although multi-dimensional stratification can improve the homogeneity of data, but it neglects the differences in the importance of samples within each layer. In practical applications, low-frequency critical events such as "energy consumption peaks" may still lack sufficient sample points under proportional sampling, resulting in insufficient learning of the model. Hence, to address this, a weighted sampling method for critical events was introduced by the researchers to enhance the representativeness of rare critical samples artificially. The 'critical event' in our combined process air conditioning system research is clearly defined as operational state transition points that result in drastic or abnormal shifts in system energy consumption. Set threshold of energy consumption change rate and extreme weather, we got about 150 critical event samples, mainly including: (1) Production Load Changes:Batch based production processes create irregular switches between active/inactive production loads in some areas; (2) Seasonal Operational Mode Transition Days: Apart from winter/summer mode switching there are cyclical seasonal changes where the system cycles through its various modes; (3) Abnormal

Weather Response Events: For example sudden summer afternoons with unexpected thunderstorms cause outdoor humidity to suddenly increase, causing dehumidifier demand to surge or extremely cold winters with sudden cold waves dropping humidity levels rapidly increasing humidifier demand.(4) Equipment restarts following large-scale overhauls - for example chillers switching from being off to full load operation on the first hot day of summer. To measure the effectiveness of different sampling methods to capture these rare samples, a "Critical Event Coverage Rate" is introduced, which is calculated as shown in equation (3):

$$R_c = \frac{S_k}{T_k} \times 100\% \tag{3}$$

In Equation (S_k S_k 3), represents the number of key events contained T_k in T_k the extracted training samples, while represents the total number of key events in the dataset. The schematic diagram illustrating the weighted sampling of key events is shown in Figure 2.

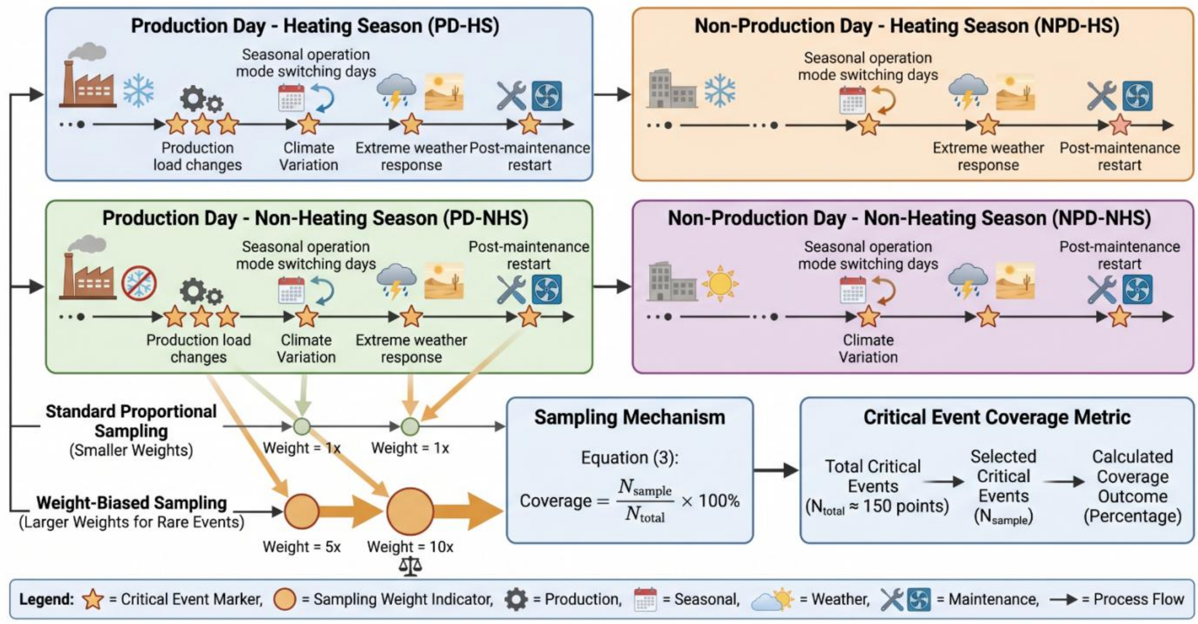


Figure 2: Schematic diagram of weighted sampling for key event weights

Figure 2: First, we do event feature extraction on the input single-layer combined data (such as "Production Day - Heating Season" layer), calculate some indicators like instant power consumption and energy consumption change rate for decision-making. Then each sample will be identified and classified according to a set of pre-set rules, so that the data stream can be divided into a "Key Event Sample Subset" and a "Conventional Sample Subset". And for rare but very important key event samples, use weighted or over-sampling method to increase their weight in sampling. For normal samples, standard proportional sampling is still used. Lastly, combine the results from both subsets of samples into one last sampled output which provides an excellent quality sample base. The quantified weight bias mechanism is presented in Equation (4) along with an updated formula for calculating intra-layer sample count.

$$\dot{n}_{\mathrm{h}} = n \cdot \frac{N_{\mathrm{h},ce} \cdot w_{ce} + N_{\mathrm{h},normal} \cdot w_{normal}}{\sum_{i=1}^{L} (N_{i,ce} \cdot w_{ce} + N_{i,normal} \cdot w_{normal})} \tag{4}$$

In Equation ($N_{h,ce} 4 N_{h,ce}$), represents the sample w_{ce} size of w_{ce} key events $N_{h,normal}$, while represents $\hbar$the weight w_{normal} of key event w_{normal} samples.The number of ordinary samples in the layer is represented by it, and the weight of ordinary samples is represented by it. Study integrate 2 core methods-Multidimensional Combinatorial Stratification and Key Events Weight Bias, form a whole, more refined level-wise sampling method, specific work flow is as follows figure 3 shows.

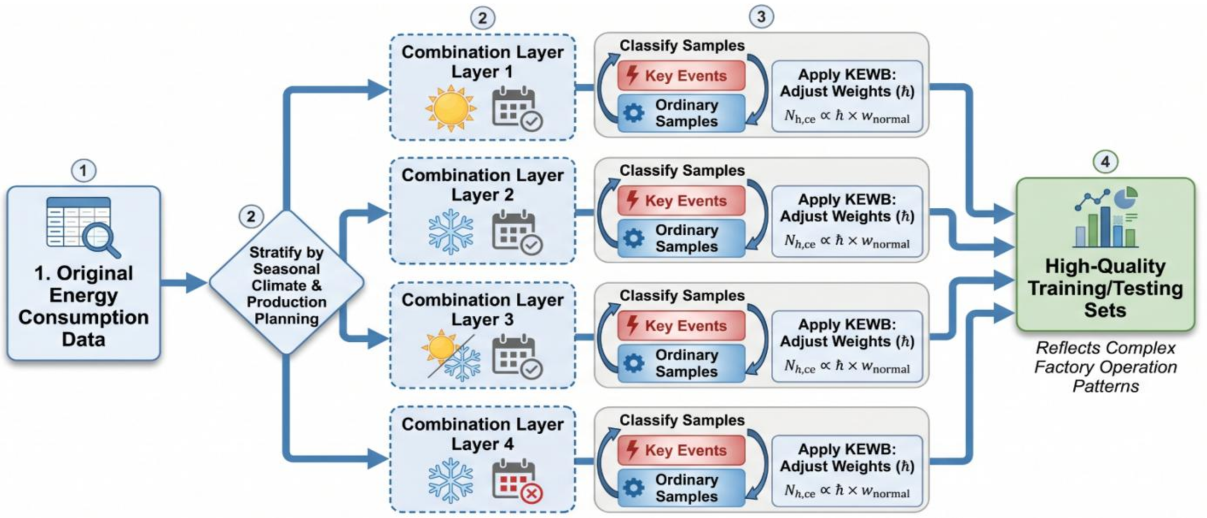


Figure 3: Flowchart of the improved hierarchical sampling method

In Figure 3, it starts from the original energy consumption data. First, it is divided again by "seasonal climate" and "production planning", breaking down the dataset into four fine combination layers. Weighted sampling is then cyclically applied to each combination layer. To extract temporal features and detect key events, samples in every layer are classified and given different kinds of samplings. Finally, the sampled results are gathered together for a high quality training set and testing set which reflects the complicated operation pattern of a cigarette factory. In short, this paper puts forward an improved progressive hierarchical sampling method. "Multi-dimensional Combination Layering" resolves the problem of coarse data patterns, "Key Event Weight Bias" solves the problem of insufficient learning of rare but important samples to improve sample quality.

2.2 Construction of the extreme random forest prediction model

Through the improvement of stratified sampling methods, a high-quality dataset has been formed that fully demonstrates the complicated operational patterns of tobacco factories. If this kind of data can't help in forecasting the energy needs for the next few days, the company's planning about when to use energy will still have to be based on old historical information. This could lead to "oversupply" and waste of energy, or "under-supply" which would affect normal production plans, thus damaging business profits [15, 16]. So developing a new prediction model is important if we want to do better energy management. The traditional statistical forecasting models cannot reflect the complex nonlinear relationship between the energy consumption of the tobacco factory, and single-machine learning model such as decision tree has weak generalization ability. For this reason, the study takes RF as the main forecasting model. RF improves the stability of the model by combining the outcomes of various decision trees, with the basic structure depicted in Figure 4.

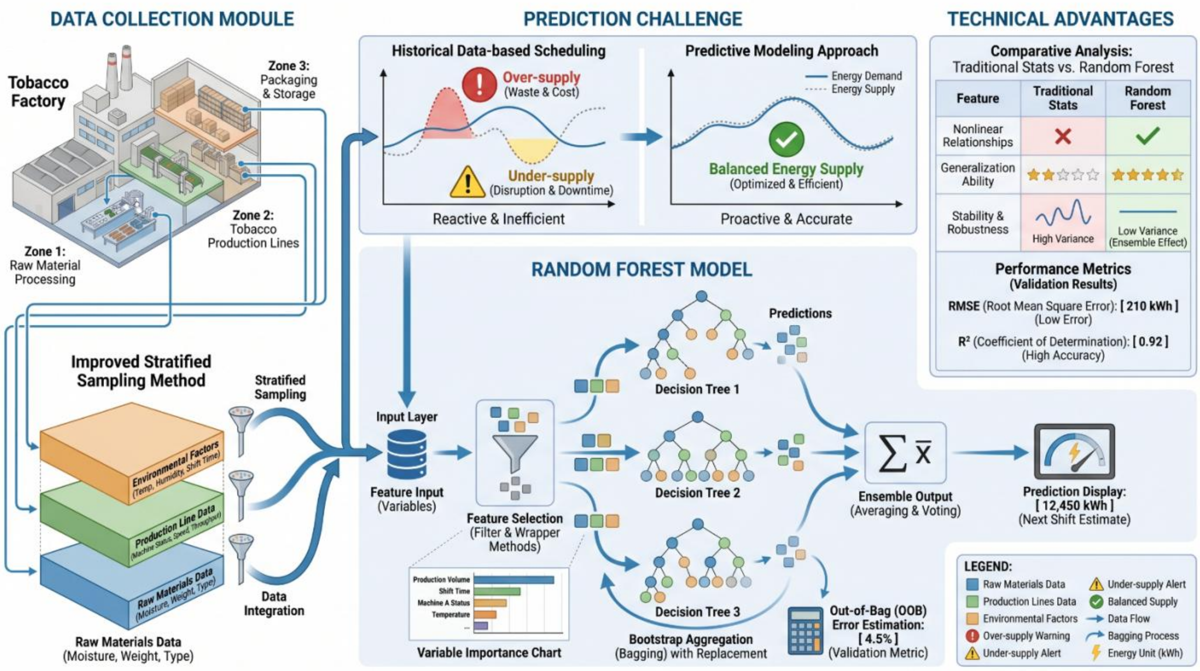


Figure 4: Schematic diagram of the Random Forest algorithm architecture

Figure 4 shows how multiple different training sample subsets are produced by a high-quality dataset that has been processed through improved hierarchical sampling with random sampling. Then, one decision tree is independently trained on each of the samples to get m trees in the forest. In the prediction phase, each tree produces an independent prediction for the input data. And finally, the last forecast is made by combining all these forecasts together. This ensemble strategy's effectiveness can be analyzed theoretically using its upper bound of generalization error as described in Equation (5) [17].

$$PE^* \leq \frac{\bar{\rho}(1-s^2)}{s^2} \tag{5}$$

In Equation (5), $\bar{\rho}$ represents the average correlation of prediction results ∇between any two decision ∇trees∇, while represents the average strength of a single decision tree. The expression for is given in Equation (6).

$$s = E_{X,Y} mg(X,Y) \tag{6}$$

In Equation ($\boldsymbol{E}_{X,Y}$), represents Xthe expected Yvalue Yof the Y.Energy consumption data distribution and its corresponding real energy consumption values. Random sampling of RF will be diverse but equal-probability sampling may pick features with very weak correlation, making model not robust. In order to solve the problem, this paper improves the candidate feature generation in node splitting by using a weighted feature selection method. This strategy can replace equal probability sampling to guide the model to select the main features and avoid the blindness of completely random selection. The probability expression for selecting a feature is shown in equation (7).

$$P(A_i) = \frac{w_i}{\sum_{j=1}^{M} w_j} \tag{7}$$

In Equation (5 $\bar{\rho}$ $\bar{\rho}$), represents the average correlation of prediction results s s between any two decision s trees s, while represents the average strength of a single decision tree. The expression for is given in Equation (6).

$$s = E_{X,Y} mg(X,Y) \tag{6}$$

In Equation ($E_{X,Y}$), represents X the expected X value Y of the Y distribution of energy consumption data and its corresponding actual energy consumption values. The random sampling s employed s by the RF algorithm ensures diversity; however, equal-probability sampling may select weakly correlated features, compromising model robustness. In order to solve this problem, the study improves candidate feature generation during node splitting by adding a weighted feature selection strategy. This strategy replaces equal-probability sampling, leading the model to focus on important features, reducing the blindness of purely random choices. The probability expression of feature selection is shown in equation (7).

$$P(A_i)=\frac{w_i}{\sum_{j=1}^{M} w_j} \tag{7}$$

Equation (7)P(Ai), in which Aiprobability Ai denotes the probability that the i-th feature in the dataset i is selected M as a candidate M; it represents the global importance weight of the i-th s feature; and it denotes the s total number of features.

Weighted feature selection strategy can greatly improve the performance of the model by improving the candidate feature set. but once you've done your features pickings, then RF will need to go through all those possible cuts again just like before and that would take way too much computer time for it not even knowing about anything else than what he had been taught from training data sets right now so basically it is going to learn more things as well. The study proposed a segmentation point randomization mechanism to construct the ERT model. this replaces expensive optimal search with a segmentation point randomization technique that decreases overfitting and increases training efficiency. Figure 5 shows how the ERT model differs from the RF algorithm.

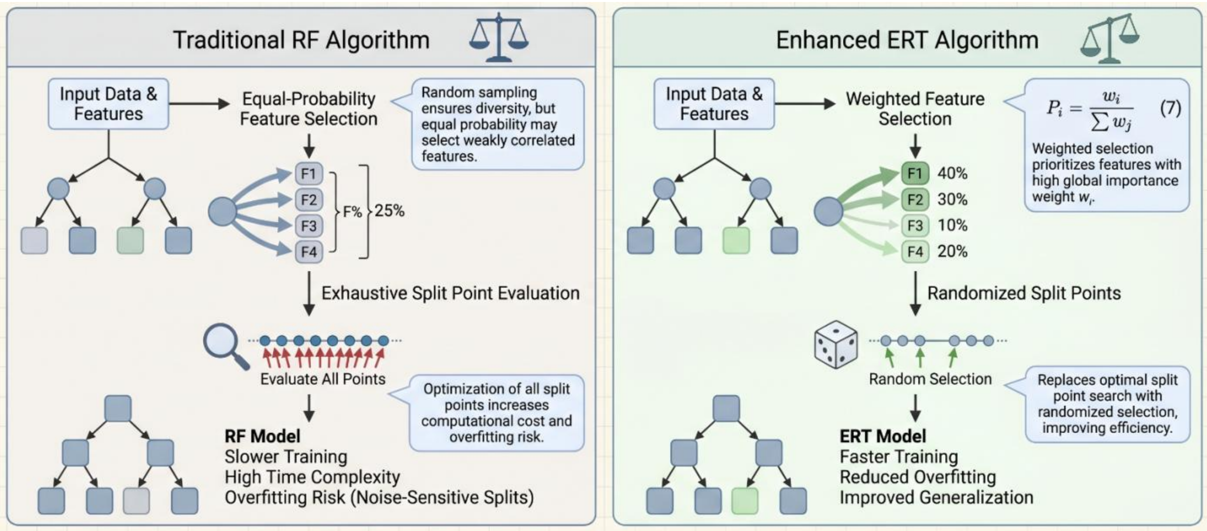


Figure 5: Comparison of RF and ERT node splitting mechanisms

In Figure 5(a), the RF segmentation mechanism iterates through each candidate feature and all segmentation points, calculates the information gain to find the locally optimal segmentation point, then selects the globally optimal solution from these local optima for node splitting. In figure 5(b) ERT iterate K times by randomly selecting features and split point to get k random candidates. Finally,ERT chooses the solution with the maximum information gain out of these k solutions to carry out node division. Information gain formula is shown in Equation (8)[18].

$$G(D,a) = H(D) - H(D \mid a) \tag{8}$$

In Equation (G （D,a）), represents D_a the information H（D）gain H（D）of D the feature H（$D|a$）set D on a the H（$D|a$） dataset; represents a the information D Entropy of dataset before partitioning; and is the conditional entropy of the dataset after feature based partitioning. This paper integrates the process of data processing and model construction into one stage, eventually forms an energy consumption prediction model that is based on improved hierarchical sampling and ERT, as a whole, it is shown in figure 6.

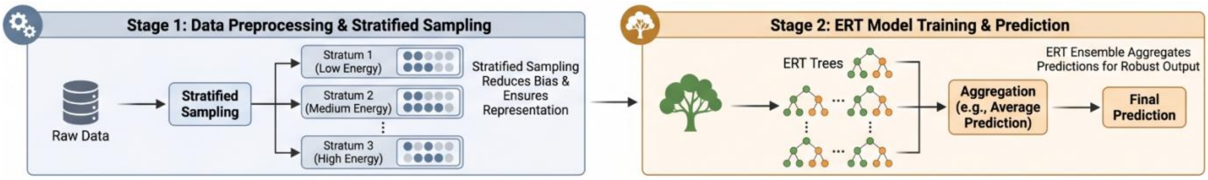


Figure 6: Framework diagram of the energy consumption prediction model based on improved hierarchical sampling and ERT

In Figure 6, the process stage of data processing is carried out. The "improved hierarchical sampling" technique was employed during this phase to create superior quality training and testing datasets from the original tobacco factory data. Model Construction and Prediction Phase uses these High-quality Training Datasets for ERT Modeling. After the training has been finished, we can use the resulting ERT prediction model on new data to produce final forecasts about energy consumption. In summary, an improved hierarchical sampling method at the level of data that integrates multi-dimensional combinatory stratification and key event weighting is established. At the level of models, it first introduces weighted feature selection with optimized node splitting, then a randomized segmentation point mechanism is introduced in order to eventually build up an extreme random forest model. The combination of the two parts produces a prediction model specifically made for forecasting energy consumption in tobacco factories.

3 Energy consumption prediction model experimental analysis

3.1 Improved stratified sampling method validation

In order to verify that the proposed prediction model can accurately describe the energy consumption pattern of tobacco factory under complex production conditions, a series of comparative experiments are carried out. Experimental data came from an energy management system in a cigarette factory and included full year data for January 2025 to December 2025. It's important to note that the study focused on all air conditioning units (a total of 10) in the entire cigarette packaging area, which is considered as the main unit of workshop-level energy consumption, not the overall plant-level energy consumption. Thus, energy consumption levels represent workshop-level features (measured data show typical single-shift consumption between 8,000 and 9,000 kWh), consistent with operational readings of this multi-unit cluster under partial load conditions. To deal with the problem of different frequencies of source data and coarse "macro-data", we standardized all the data as "8 hours of production shift", aligning the time when production schedule, weather record and energy consumption were recorded according to timestamp. The original dataset contained 365 days × 3 shifts/day = 1,095 samples; but after minor missing values were cleaned up (~5 samples), only 1,090 complete and representative shift samples were selected for use in experimental datasets. For feature dimension differences, the minimum-maximum normalization method is used to convert all the data to [0,1]. In the process of doing the work related to the feature engineering part of the project it would be very useful if one could take care about the redundancy variables through Variance inflation factor i.e VIF>10 so that the variable do not create any kind of problem while performing the training activity on these various multidimensional features like Temperature & Humidity Metrics etc. Final list of input variables for the model is shown in Table 1.

Table 1: List of Input Variables for the Energy Consumption Prediction Model

dimension	Variable Name	Description/Unit	remarks
program of production	Production Status	0-Production halted, 1-Production ongoing	Core Classification Feature
	Planned Production Volume	Ten thousand units per batch	Directly related to production load
meteorologic condition	Outdoor dry-bulb temperature	°C	Influence on heating and cooling loads
	outside relative humidity	%	Affects the humidity load (critical)
Historical Timeline	Previous energy consumption	kW·h	Previous moment energy consumption value
Time Feature	Historical average for the same period	kW·h	Periodic characteristics
	Workdays/Holiday Days	0/1 Variable	Differentiate production scheduling modes

Specifically, the energy consumption data of a combined process air conditioning system is studied, the main collected variables are chilled water temperature, steam pressure/temperature, fan frequency and power, chiller/heater/humidifier valve opening degree, ambient temperature and humidity, set temperature and humidity, and outdoor temperature and humidity. In order to guarantee authenticity and granularity of the data we have chosen ZK series combination type process Air conditioning unit group from cigarette factory packaging workshop(production area) as our major source of data for this study,this group contains ten models which are ZK-100000. The system needs to maintain constant temperature and humidity (24±2°C, 58±5%) all year round, with each unit having a rated air volume of 100,000 m³/h. Data collection covers both equipment operation parameters and environmental conditions, see Table 2.

Table 2: Key Operating Parameters and Equipment Configuration of the Process Air Conditioning System

class	Parameter Name	Number/Range	remarks
Partition Dimension	overlay area	Production Area – Roll-Sealing and Packaging Workshop	Different from the office area and the storage area
Time Dimension (Time-of-Use Electricity Rates)	rush hour	9:00~11:00,19:00~21:00	Peak electricity price period
	During regular hours	7:00~9:00,11:00~19:00	Normal production period
	Low Peak Period	23:00 to 07:00 the next day	Nighttime heating/shift duty period
plant parameter	Fans air supply frequency	30~50Hz	Frequency Conversion Control
	Surface cooler/heater/humidifier	0~100%	valve opening
Environment Settings	Workshop temperature setting	24℃	Summer/Transition Season
	Set the humidity in the workshop	58%	Process Standard Requirements

In terms of data structure, the research has made some changes to meet the needs of practical energy management by adopting time-based and area-based analysis. Zone data was obtained from an independent metering system in the rolling packaging workshop (production zone) which mainly collects energy consumption data from the core production area, i.e., the most significant change occurs here with respect to both timing as well as how it relates directly back into actual schedules related solely towards what we call "core" within such a given situation but not so much anything else like maybe even just plain old office stuff plus storage places etcetera too! Time categorization is performed according to peak and off-peak electricity pricing times, such as peak hours 9:00–11:00; Off-Peak Hours 23:00–7:00 Next Day. From the statistical point of view, during peak hours in the production area the average load of the air conditioning system is about 1.8 times that of non-peak hours, further confirming the necessity for improved hierarchical sampling. Tobacco's conditions are complicated therefore I divide my dataset as 6:4 where there are 654 training samples and 436 test samples. The hardware configurations and key parameters used for the experiments are shown in Table 3.

Table 3: Hardware environment and key parameters of the experiment

project	configure
operating system	Windows 11 64-bit Professional Edition
internal storage	32 GB
central processing unit	Intel Core i7-12700H
graphics processing unit	NVIDIA GeForce RTX 3060
programming language	Python 3.9
Main Framework/Library	Scikit-learn, Pandas, NumPy, Matplotlib
Deep Learning Framework	PyTorch 1.10
Number of decision trees	100
Random segmentation point count (K)	10(ERT model)
Maximum depth	6(XGBoost/GBDT),None(RF/ERT)
learning rate	0.1(XGBoost/GBDT)
Minimum sample size for splitting	2
Subsampling Rate	0.8

The study conducts comparative experiments for the impact of sample size on model performance through simulation and extraction of different sample sizes from a total of 1090 samples to be tested. These are displayed in Figure 7. In fig. 7(a), the bias rate of improved hierarchical sampling is only 0.9%, using the data set of 1,090 samples, which is significantly less than that of one-dimensional hierarchical sampling (3.2%), systematic sampling (6.2%), simple random sampling (7%). Fig. 7(b) shows that as far as the core metric of critical events coverage is concerned.

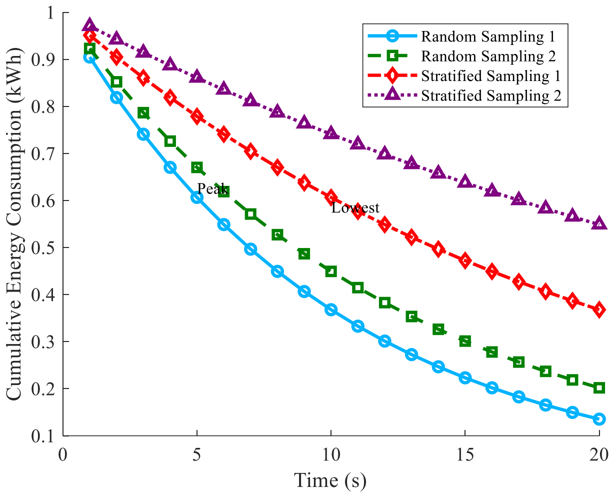


Figure 7: Performance evaluation of different sampling methods under varying sample sizes across datasets

Study goes one step further to compare the performance of different sampling methods according to the perspective of information fidelity. Comparing the four approaches according to two indicators, the extreme value retention rate and relative entropy. The result is shown in figure 8. As Figure 8(a) illustrates, if there are only 218 data samples in the dataset, then improved hierarchical sampling will have an 88.5% extreme value preservation rate while one-dimensional hierarchical sampling has a 36.8%.

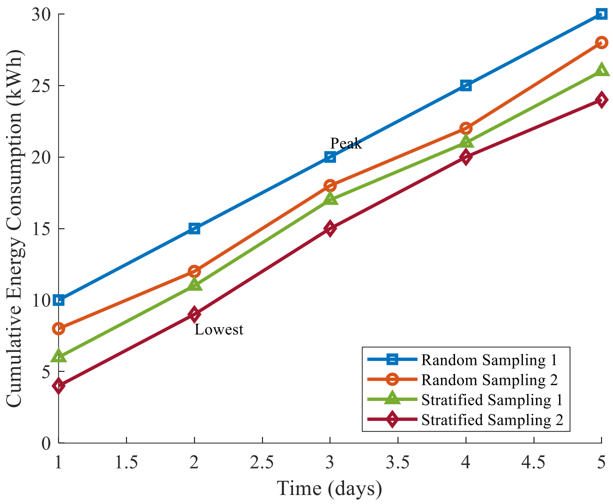


Figure 8: Comparison of performance of different sampling methods in terms of information fidelity

3.2 Comparison of ERT Core Prediction Model Performance

In order to assess the overall performance of ERT in real energy scheduling for tobacco enterprises, we compared XGBoost, Gradient Boosting Decision Tree (GBDT) and Random Forest (RF) model. Two typical operating scenarios are selected for experimentation that directly address core energy management issues. Results were assessed using two metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as shown in Figure 9. As seen in Figure 9a, in the "production day-heating season" scenario, the RMSE of the ERT model was 85.3kW·h, which is less than the RMSE of the XGBoost model at 88.1kW·h, while the MAE of the ERT model was only 61.5kW·h, the lowest among all models. In the relatively simpler "non-production day–non-heating season" case shown in Figure 9(b), the ERT model obtained RMSE and MAE values of 29.5 kW·h and 18.1 kW·h respectively, a decrease of 4.93% and 5.82% compared to the RF model’s 31.2 kW·h RMSE and 19.5 kW·h MAE. These findings show that the ERT model has less prediction error and improved robustness across different complexity levels of operational situations.

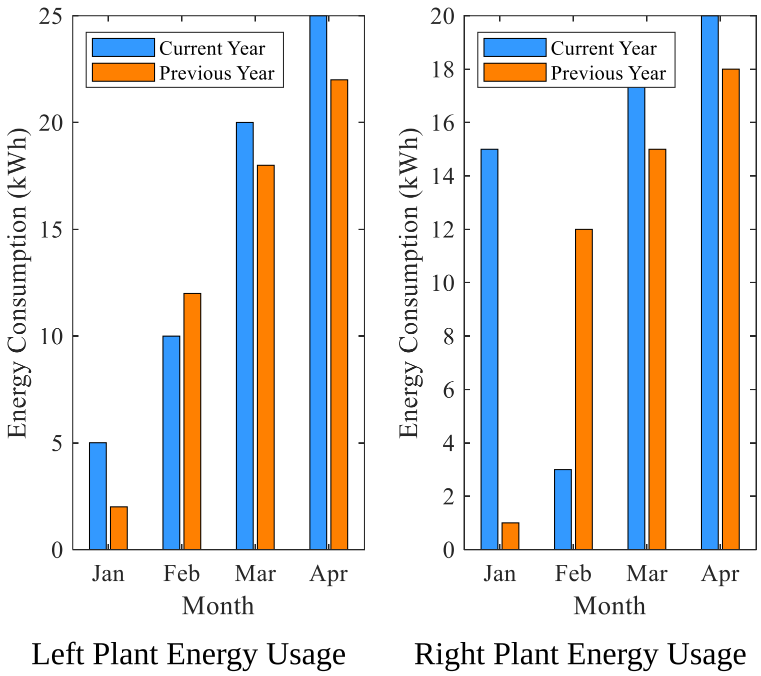
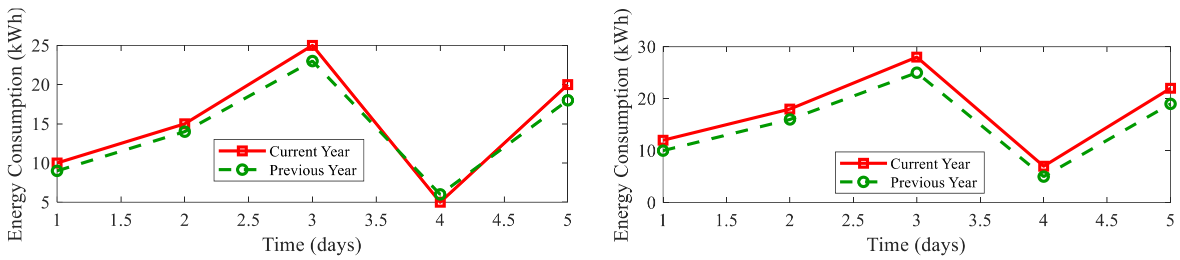


Figure 9: Comparison of performance of different prediction models under typical conditions

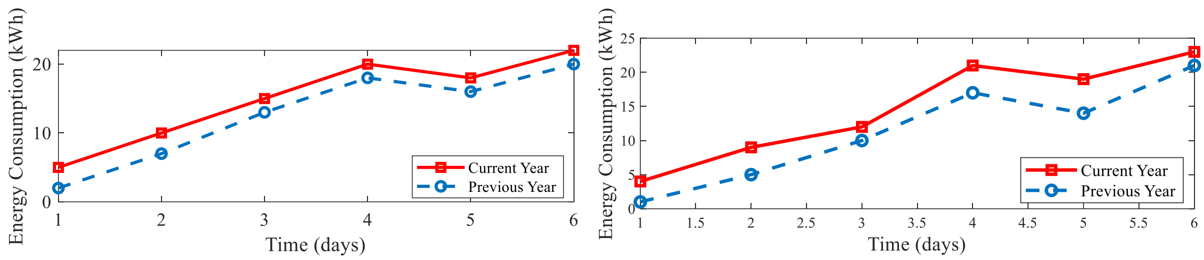
The study picked a 30-day non-stop timeframe from the test dataset including weekends and busy days as an experiment situation, compare the prediction curve of different models with actual data curve. As shown in Figure 10, as seen from figure 10(a), the ERT model showed the best accuracy for the difference between its prediction and the real value during the energy consumption trough around day 10 with a predicted value of approximately 5,850 kW·h vs an actual value of 5,900 kW·h. Figure 10(b) shows that although XGBoost model basically captured the overall fluctuation trend, large deviation appeared between its predicted result and real value. Figure 10(c): GBDT model's prediction value is much lower than the actual value close to the peak on day 15. Figure 10(d): RF model had obvious lagging, it was only about 5650Kw.h at the bottom on day25, far from the real number. This demonstrates that the ERT model is better at identifying important transition points like changes in production status (e.g., shutdowns on weekends or returning after holidays).



(a)Left Plant Energy Profile (b)Right Plant Energy Profile

Figure 10: Comparison of energy consumption prediction curves of different models

In order to demonstrate more effectively that the proposed model is effective and reliable for practical engineering problems, we did a thorough case study of the modular process air conditioning system at the production core. Dataset - it has all energy consumption data of process air conditioning system from 4th april (Friday) to 8th April(Tuesday), 2025 as test dataset. It represents a period which is extremely representative that includes 3 typical operation mode "Friday Production Day – Spring Operating Condition", "Weekend Non-Production Day – Off Peak Period"," Monday Production Day – Resumption". The biggest problem is whether or not the model can correctly detect this very important "state change" when going from off during weekends to on during Mondays. Input features are operation parameters of air conditioner system plus real-time meteorological data obtained via aforementioned improved sampling technique. Figure 11 compares the prediction curves of each model with the actual energy consumption curves over these five days. And the ERT model's predicted curve most closely matches the true values, especially at key time points. As shown in Figure 11, the ERT model captured the step-by-step increase in load after resuming operations on Monday April 7th. At the peak hour of 2pm on April 7th, the actual usage was 224kWh and ERT predicted 220kWh with very little error. Conversely, XGBoost has large delay to react on Apr7's Resumption, GBDT under-predicted afternoon peak quite badly too. RF Model shows obvious Overfitting fluctuation on Weekend Low Load Period



(a)Plant A Energy Profile (b)Plant B Energy Profile

Figure 11: Comparison of sub-models for five-day energy consumption forecasting of process

To further quantitatively study the model's performance in different modes of operation, the statistical analysis of prediction errors was performed and the economic dispatch value was evaluated together with local time-of-use electricity pricing (peak period: 9:00-11:00). See Table 4 for the specific result. As can be seen from the table, under the complex "Production day - Heating Season" condition, the improved sampling method allows the model to learn effectively from winter-related important events (like strong cold fronts and low temperatures) to keep a very high level of accuracy at MAPE <6.1%. The MAPE decreased to 1.7% during relatively stable "Non-production day - Non-heating season" conditions, showing strong generalization ability. According to the accurate prediction of the model, it can generate an operable dispatch instruction, such as when forecasting "there will be a very cold wind tomorrow", the dispatch center could increase the steam supply pressure by one hour before to improve the humidification capacity. This way we not only avoid severe load fluctuations on the main unit but also achieve cost savings through TOU pricing. this case show that model is developed have large engineering potential turn data into actionable energy saving strategy.

Table 4: Prediction errors of the ERT model for different operating modes of the process air conditioning system

Operating Condition Combination	RMSE(kW)	MAE(kW)	MAPE(%)	Key Challenges and Model Performance
Production Day-Heating Season	85.3	61.5	6.1	The system successfully detected load fluctuations caused by license plate changes and sudden weather conditions, achieving accurate peak load forecasting.
Non-production Days – Non-heating Season	29.5	18.1	1.7	Accurately identify the baseline of the system operating under low load; the prediction curve remains stable without any abnormal fluctuations.

In order to clarify the physical mechanisms behind the ERT model accurately capturing energy consumption troughs and sudden operational changes, it is extracted the feature importance of each operation mode as shown in Figure 12. Figure 12(a) shows that "outdoor relative humidity" reached a maximum score of 0.42 under "production mode," followed by "planned output volume." It meets the strict requirements for the humidity environment in the tobacco packaging workshop (about 58%±5%) to show that energy consumption mainly comes from the humidity load. especially the difference between outdoor RH and your workshop’s setting, if you’re out there - high powered deep cooling/drying or adding moisture either way both take large amounts of energy. Figure 12(b) indicates that when the machine is not running (non-production time), the "Production status" becomes important at 0.45, and the importance value of "Preoperative power consumption" also increases, and the importance value of "Outdoor dry bulb temperature" also increases. This means during shutdown periods, the model will reduce its predictions down to a basic level by utilizing "shutdown" signals but it would focus on thermal inertia effects. The condition-based adaptive feature weighting mechanism could account well for the performance differences shown in figure 10:RF models are too dependent on global smoothing features, while ERT model correctly detects abrupt switching from "production status" to "humidity fluctuations", allowing quick response to the low point at Day 5 without suffering from lagging and overestimation commonly seen in traditional models.

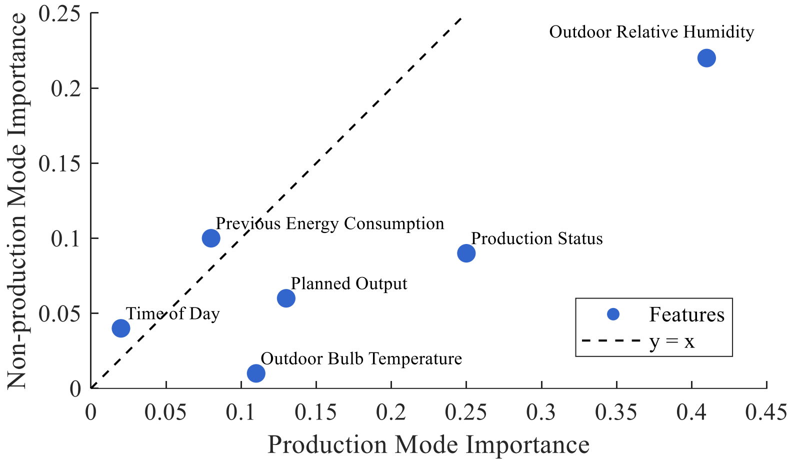


Figure 12: Comparative analysis of feature importance under different operating modes

3.3 Integrated Prediction Model's Comprehensive Performance Analysis

To show how all the modules of the integrated model work together to tackle the particular issue of energy usage prediction in tobacco factories, the research has set up an ablation test which can be found in Table 5. As shown by Table 5, when "Multidimensional combination hierarchical"(Improvement 1) was introduced alone, it increased from 0.79 to 0.86 for the model's determination coefficient (R^2). The RMSE value is 131.5 kWh if you only add the "Key Event Weight Tilt" Module (Improvement 2). Combining two data processing modules (Improvement 3) raises R^2 to 0.91. On top of Improvement 3 we added "Weighted Feature Selection"(Improvement 4), this further raised R^2 to 0.93. Final model with all improvements (Improvement 5) reaches R^2=0.95, and reduces both rmse and mae to 44.1kwh and 25.2kwh respectively. In terms of computational efficiency, the segmentation point randomization mechanism (Improvement 5) implemented significantly shortened training time: the final model trained in just 5.9 seconds, a 58% speed-up over the baseline random forest model's 14.1 seconds. This comes about as the ERT algorithm does away with that very intensive "Feature Sorting" & "Optimal Segmentation Point Traversal" during Node Splitting and substitutes that with a Random Partitioning Strategy which speeds up Training at the Algorithmic Complexity Level. It proves that not only did the 'Multi Dimensional Stratification' take care of combined Production Seasonality, but also ensured that none of those important production points were missed due to 'Bias Weights'. Both these things improved the accuracy of the model in predicting energy consumption at Tobacco Factories under complex conditions.

Table 5: Ablation experiments of each improved module

model	+Multi-dimensional combination and layering	+Key Event Weight Bias	+Weighted Feature Selection	Randomize division points	RMSE/ kW h	MAE/ kW h	R^2	training time /s
stratified sampling + random forest	×	×	×	×	145.2	80.1	0.79	14.1
Improvement 1	√	×	×	×	118.0	65.3	0.86	13.9
Improvement 2	×	√	×	×	131.5	72.4	0.82	14.3
Improvement 3	√	√	×	×	90.2	50.5	0.91	13.6
Improvement 4	√	√	√	×	82.8	46.1	0.93	13.1
Improvement 5	√	√	√	√	44.1	25.2	0.95	5.9

Improved Hierarchical Sampling, Improved ERT, Model 2: Combines traditional sampling with gradient boosting decision trees; High performance model integrating traditional sampling with extreme gradient boosting trees; Benchmark model using traditional sampling and classic random forests (Model 4). Under the two typical working conditions, we assess the coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE) for all models, with the results displayed in Figure 13. As shown in Fig. 13a, Model 1 has the highest $R^2 = 0.91$, MAPE = 6.1%. Model 3 ranks second with R^2=0.89 and MAPE=6.5%. In figure 13b, Model 1 achieves an R square of 0.975, MAPE is 1.7%, while model 4 obtains an R square of 0.965, MAPE is 1.8%. It is evident from the above that Model 1 can give good technical help to tobacco enterprises on precision energy management and cost optimization for complex production and seasonal load change problems.

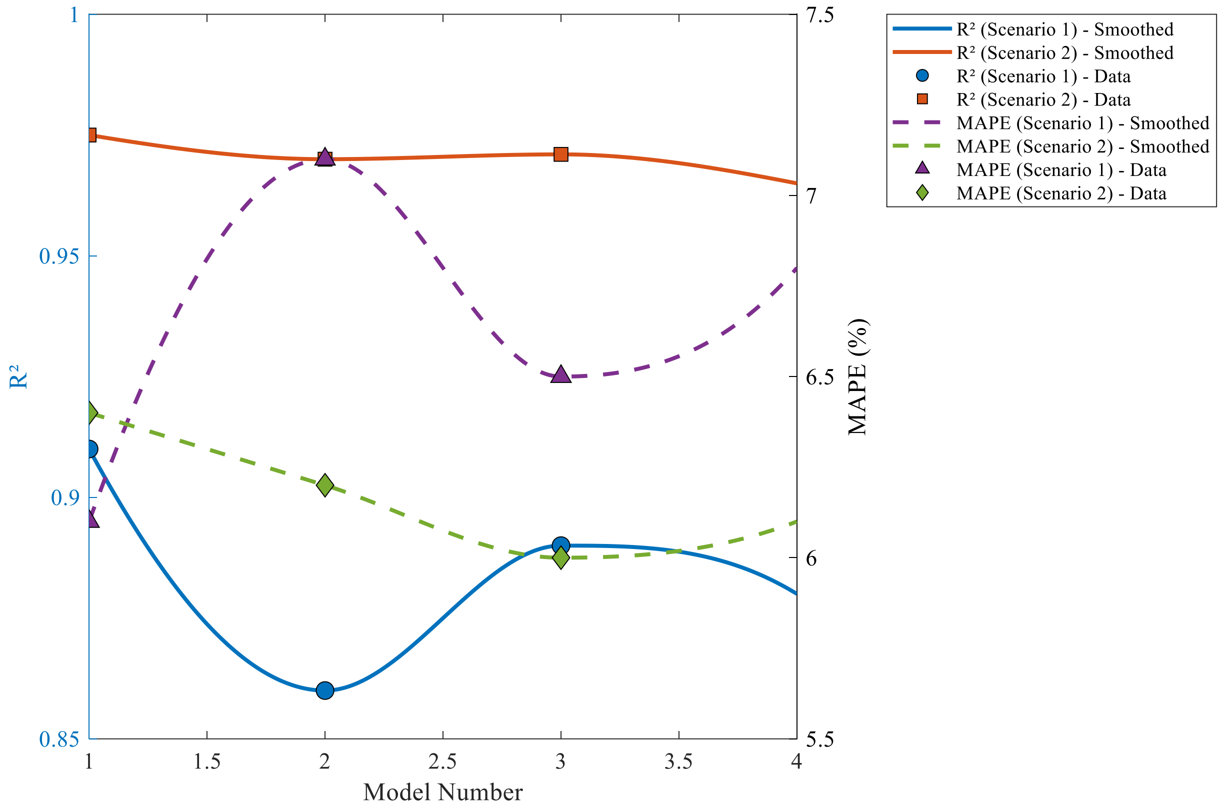


Figure 13: Comparison of R^2 and MAPE metrics across different operating conditions for each model

4 Conclusion

Energy consumption of tobacco enterprises is a complex feature with overlapping multimodal characteristics and nonlinear fluctuations, making the traditional forecasting model difficult to achieve high precision. Based on these characteristics, this paper proposes an enhanced energy consumption prediction model integrating improved hierarchical sampling and Event-Related Theory (ERT). Taking air-conditioning system in tobacco factory as case which belong to kind of coupling multi-mode coupled energy consumption system, firstly we propose a improvement of hierarchic sampling method by means of multidimensional combinatorial stratification and key-event weighting, then an extreme random forest model incorporating weighted features and randomized segmentation points is adopted. Experiments show that the deviation rate is only 0.9% when the sample size is 1,090.

(1) On "production day-heating season" in a complicated working environment, the RMSE of ERT model is 85.3 kW·h and MAE is 61.5 kW·h. Case analysis such as resumption of work on weekends confirms that it can accurately predict load fluctuations caused by changes in production schedules like grade switching or external weather events like sudden thunderstorms, achieving a MAPE below 6.1% under normal production conditions, providing practical advice for proactive energy planning.

(2) Ablation experiments show that every suggested enhancement significantly improves the model's performance, giving R^2=0.95. "Non-production day - non-heating season" MAPE stays at 1.7%. In short, the combined forecasting model, through a synergistic improvement of data processing and algorithm optimization, solves the problem of predicting energy consumption in the complex working conditions of tobacco factories, providing reliable technical support for more accurate management of energy.

(3) But it does depend on having a definition given beforehand of what the "key events" in tobacco factories are, and finding the correct hyperparameters for the models still needs to be trial and error. Future research direction: Add adaptive recognition mechanism of key events, so that the model is smarter and can have more application areas.

About the Author

Jianqi Yin (1995-), female, from Jining City, Shandong, graduated from the School of Mechatronic Engineering of Shandong University of Science and Technology with a master's degree, engaged in research on electrical control and intelligent manufacturing.

Shilai Yuan (1989-), male, from Shaoxing City, Zhejiang, graduated from the School of Computer Science of University of Nottingham with a master 's degree, engaged in research on intelligent manufacturing and green low-carbon technologies.

MinJie Zhu(1991-), male, from Hangzhou City, Zhejiang, graduated from the School of Internet of Things of Jiangnan University with a bachelor's degree, engaged in research on energy management and electrical control.

References

[1] Cai Xia, Wang Ke, Zheng Junhong, et al.(2024). Research on Intelligent Supply and Control of Power and Energy for Manufacturing Enterprises [J]. Journal of Engineering Design, 31(6):697-706.

[2] Cai Xia, Wang Ke, Zheng Junhong, Cai Xia, Wang Ke, Zheng Junhong, et al.(2024).Research on intelligent supply and management of power energy for manufacturing enterprises[J]. Chinese Journal of Engineering Design, 31(6):697-706.

[3] Huang Kehua, He Junfei, Jia Chunsheng, et al.(2025). Research and Practice on Digital Technology Assisting Green Spinning Manufacturing [J]. Cotton Textile Technology, 53(11):45-51.

[4] Huang Kehua, He Junfei, Jia Chunsheng, Huang Kehua, He Junfei, Jia Chunsheng, et al.(2025). Research and practice of digital technology facilitating green manufacturing in spinning[J]. Cotton Textile Technology, 53(11):45-51.

[5] Zhang C, Chen Y, Chen H, [3]Zhang C, Chen Y, Chen H, et al.(2024). Industry 4.0 and its implementation: A review[J]. Information Systems Frontiers, 26(5): 1773-1783.

[6] Wu Jinlu, Hu Anfu, Jiang Jian, et al. (2025). Research Progress on the Generation and Migration Release Patterns of Moisture in Heated Cigarettes [J]. Henan Agricultural Sciences, 54(8):15-25.

[7] Wu Jinlu, Hu Anfu, Jiang Jian, Wu Jinlu, Hu Anfu, Jiang Jian, et al.(2025). Research Progress on Generation,Migration and Release of Moisture in Heated Tobacco Product[J]. Journal of Henan Agricultural Sciences, 54(8):15-25.

[8] Huang Changmin, Ye Min, Huang Bo.(2025).Research on Enhancing Digitalization Capabilities of Cigarette Factory Equipment Based on the Overall Equipment Effectiveness (OEE) Theory [J]. Automation and Instrumentation,(10):282-286.

[9] Huang Changmin, Ye Min, Huang Bo.(2025). Research on Enhancing the Digital Capability of Cigarette Factory Equipment Based on the OEE Theory of Equipment Comprehensive Efficiency[J]. Automation & Instrumentation,(10):282-286.

[10] Chen Qizhen, Wang Xu, Jiang Chuanwen, et al.(2025). Research on operational strategies for switching substations considering seasonal complementarity of adjustable capacities in energy-backup markets [J]. Grid Technology,49(3):1056–1069.

[11] Chen Qizhen, Wang Xu, Jiang Chuanwen, Chen Qizhen, Wang Xu, Jiang Chuanwen, et al.(2025). Operational Strategy for the Participation of Battery Swapping/Charging Stations in the Energy-reserve Market Considering Seasonal Complementarity of Adjustable Capacities[J]. Power System Technology,49(3):1056-1069.

[12] Zuo Hao, Wang Yulin, Jin Rui, et al.(2022). Machine learning-based energy consumption modeling method for hydraulic drive units [J]. Journal of Hefei University of Technology (Natural Sciences Edition),45(5):582–588,619.

[13] Zuo Hao, Wang Yulin, Jin Rui, Zuo Hao, Wang Yulin, Jin Rui, et al.(2022). EEnergy consumption modeling method of pump unit based on machine learning[J]. Journal of Hefei University of Technology(Natural Science), 45(5): 582-588, 619.

[14] Ullah I, Liu K, Yamamoto T, [8]Ullah I, Liu K, Yamamoto T, et al..(2022). A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability[J]. Energy & Environment, 33(8): 1583-1612.

[15] Yi Wangyuan, Yin Ruixue, Tian Yingquan, et al.(2024). Research on Energy Consumption Prediction and Multi-objective Optimization of Cutting Parameters for CNC Milling [J]. Journal of Chongqing University of Technology (Natural Sciences), 38(3):240–249.

[16] Yi Wangyuan, Yin Ruixue, Tian Yingquan, Yi Wangyuan, Yin Ruixue, Tian Yingquan, et al.(2024). Research on energy consumption prediction and multi-objective optimization of cutting parameters in CNC milling[J]. Journal of Chongqing Institute of Technology, 38(3): 240-249.

[17] Tang Xishu, Tian Dexing, Fang Ruiping, et al.(2024). Research Progress on Stick Separation Technology and Equipment in Cigarette Manufacturing Process [J]. Journal of Light Industry, 39(4):109-117.

[18] Tang Xishu, Tian Dexing, Fang Ruiping, Tang Xishu, Tian Dexing, Fang Ruiping, et al.(2024). Research progress on stem sliver separation process and equipment in cigarette manufacturing[J]. Journal of Light Industry, 39(4): 109-117.

[19] Wang W, Zhao F, Wang M, [11]Wang W, Zhao F, Wang M, et al..(2025). Multi-objective optimisation of cigarette production planning and inventory management[J]. International Journal of Computing Science and Mathematics,21(1): 64-76.

[20] Zhou Ya Wen, Chen Wang Xue, Deng Cui Hong, et al.(2023). Optimal estimation of parameters under inverse exponential distribution with three sampling designs [J]. Systems Science and Mathematics,43(4):1069–1080.

[21] Zhou Yawen, Chen Wangxue, Deng Cuihong, Zhou Yawen, Chen Wangxue, Deng Cuihong, et al.(2023). Optimal Estimation of the Parameter of Inverse Exponential Distribution Under Three Sampling Designs[J]. Journal of Systems Science and Mathematical Sciences, 43(4): 1069-1080.

[22] Che Jinyong, Weng Zhenghao, Sheng Guoqing, et al.(2025). Robust separation method based on consistent support surface shape under random sampling [J]. Acta Optica Sinica,45(19):249-259.

[23] Che Jinyong, Weng Zhenghao, Sheng Guoqing, Che Jinyong, Weng Zhenghao, Sheng Guoqing, et al.(2025). Robust Support-Induced Error Separation Method Based on Random Sample Consensus[J]. Acta Optica Sinica,45(19): 249-259.

[24] Makwana D, Engineer P, Dabhi A, Makwana D, Engineer P, Dabhi A, et al.(2023). Sampling methods in research: A review[J]. International Journal of Trend in Scientific Research and Development, 7(3): 762-768.

[25] Zhang Jiyu, Gu Yajun, Zhang Qi, et al.(2025). Energy efficiency improvement and case study of heating systems under the dual-carbon framework [J]. Hebei Industrial Science and Technology, 42(3):240-247.

[26] Zhang Jiyu, Gu Yajun, Zhang Qi, Zhang Jiyu, Gu Yajun, Zhang Qi, et al.(2025). Energy efficiency improvement of heating systems under the dual carbon context and case analysis[J]. Hebei Journal of Industrial Science & Technology, 42(3): 240-247.

[27] Rasanjali W A, Mendis A, Perera B, [16]Rasanjali W A, Mendis A, Perera B, et al.(2024).Implementing enterprise resource planning for lean waste minimisation: challenges and proposed strategies[J]. Smart and Sustainable Built Environment,13(2): 330-353.

[28] Ding Sha, Shen Taorong, Zhang Yanfei, et al.(2023).Research on a tobacco extract category recognition model based on the random forest algorithm [J]. Journal of Analytical Testing,42(11):1510-1516.