An Adaptive Random Forest Model for Predicting Demands and Solar Power of a Real Integrated Energy System

Jie Mei^{1*}, Christopher H.T. Lee², James L. Kirtley¹

¹ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

² School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

*jiemei@mit.edu

Abstract: In order to address the challenges of improving energy efficiency and integration of renewable energy, multi-energy systems, composed of electric, natural gas, heat and other energy networks, have received more and more attention in recent years and have been rapidly developed. Through integration as a multi-energy system, different energy infrastructures can be scheduled and managed as one unit. One of the main stages in the optimal scheduling of a multi-energy system is the predictions of various demands and sustainable energy in the scheduling horizon. This paper proposes a prediction model based on adaptive random forest for demands and solar power of a real MES, Stone Edge Farm, in California. The adaptive random forest model can provide a probability distribution of the prediction results. This allows users to consider a variety of scenarios that may occur in the future for further system operation optimization and help users evaluate the reliability of the results. Besides, an online self-adaptability feature is implemented with the model so it can adapt to the new forecasting environment when new observations are detected. The simulations show that the adaptive random forest model is better than the benchmark models in terms of prediction accuracy.

Nomenclature

j	j th leaf node in a CART
k	Index of CART in RF model
K	Number of CARTs in RF model
m	Number of split features at each node
N	Number of samples in training data set
N^*	Number of samples in bootstrap training data set
N_R	Number of samples in received online data set
N_I	Number of neurons in input layer of NN
N_O	Number of neurons in output layer of NN
N_{HL}	Number of hidden layers of NN
N_{L1}	Number of neurons in 1 st hidden layer
N_{L2}	Number of neurons in 2 nd hidden layer
N_{L3}	Number of neurons in 3 rd hidden layer
n_{min}	Minimum decision node size
p	Number of feature variables for RF model
R_j	Received sample numbers at leaf node j
s	Split value at a decision node
S	Range of split value at a decision node
T_k	the k th CART in RF model
Z	Training data set
Z^*	Bootstrap training data set
Z_R	Received online data set
$\hat{Y}_k$	Prediction result of the k th CART
$\hat{Y}_{RF}$	Prediction result of RF model
α	Confidence interval of RF model

δ	Threshold for number of received samples
ζ	Threshold for prediction accuracy improvement
μ	Order of autoregressive part of ARMA
β	Order of moving average part of ARMA
γ_k	Prediction accuracy of the k th CART
λ	Number of repeat times for new observations

1 Introduction

In the past few decades, due to growing concerns about global warming and energy crisis, more efficient energy scheduling and management methods have been encouraged [1], [2]. In order to address the challenges of improving energy efficiency and integration of renewable energy, multi-energy systems (MESs), composed of electric, natural gas, heat and other energy networks, have received more and more attention in recent years and have been rapidly developed [3]. Traditionally, different energy infrastructures in an energy system are usually scheduled and managed independently, resulting in lower energy use efficiency and waste of resources [4]. Through integration as an MES where multiple energy carriers can be converted, adjusted, and stored to meet energy demands, different energy infrastructures can be scheduled and managed as one unit [5].

The optimal management of MES is mainly divided into two stages. The first stage is the predictions of various demands and sustainable energy in the scheduling horizon. Then the second stage is the optimal management of devices in the system based on the prediction results [6]. Many researchers have done a

series of studies on the optimal scheduling of MES, and the methods can be roughly divided into the following types: mixed integer linear programming-based methods (MILP) as in [7], [8], heuristic algorithm-based methods as in [9], [10], and stochastic optimization-based methods as in [11], [12], etc. These scheduling methods mainly use predictive models such as autoregressive moving average (ARMA) [13], neural network (NN) [14], and reinforcement deep learning (RL) [15]. However, a problem with these previous prediction models is that they cannot provide confidence intervals for predictions. When MES scheduling needs to take into account a variety of scenarios that may occur in the future, users cannot obtain the probability of these possible scenarios through these prediction models, and thus these models are no longer applicable. Besides, those models are hard to update themselves in an online manner.

This paper proposes an adaptive random forest model for the predictions of demands and solar power of a real MES, Stone Edge Farm, in California. RF models have been extensively applied in various classification and prediction applications, such as engineering [16]-[18], climate [19]-[21], transportation [22]-[24], economic [25]-[28], and medical [29]-[31]. The adaptive random forest model can provide a probability distribution of the prediction results. This allows users to consider a variety of scenarios that may occur in the future for further system operation optimization and help users evaluate the reliability of the results. Besides, an online self-adaptability feature is implemented with the model so that it can adapt to the new forecasting environment when new observations are detected. To the authors' knowledge, this is the first detailed paper describing RF applications in MES optimal scheduling problems.

The rest of this paper is organized as follows. Section 2 gives a general description of how a RF model works. Section 3 introduces the probability distribution of prediction results and the online self-adaptability function of proposed adaptive RF model. The simulation results and analysis are given in Section 4, and finally the conclusions are drawn in Section 5.

2 Introduction of RF

RF is an ensemble machine learning model for classification and prediction. It is generally based on two techniques, classification, and regression tree (CART) and Bagging. CART is a tree-structure classification model, which maps observations of items to conclusions about item categories and was first proposed by Breiman in the year of 1984. The advantage of CART is that it can fit the training data well. However, when making predictions or classifications, CART has high variance as it can easily output different results with small changes in input variables [32].

In order to solve this problem, CART was extended to RF through another important technique, Bagging, which was also introduced by Breiman in late 1996. Bagging generates more new training sets through uniformly sampling from the original training data set with replacements. The new training sets generated by Bagging can train more CARTs to reduce the variance of a single CART, while maintaining a low bias. In

addition, RF uses a technique called random node optimization to further reduce the CART variance, that is, the decision variables at a decision node of a CART will be randomly selected from a pre-determined feature variable set. All of the above-mentioned modifications to CART by RF have avoided the shortcomings of CART and proved that good prediction accuracy can be achieved.

Fig. 1 shows a simple illustration of CART and RF. From the figure, we can see that a CART makes a decision in each decision node based on the splitting of feature variables and goes all the way down until it reaches a leaf node. In this simple illustration, the CART on the top has 4 prediction results (leaves) determined by two feature variables. The first classification is determined by the splitting of the feature variable in the first row, and the second classification is determined by the splitting of the feature variables in the second row. Fig. 1 also provides a hint for the CART growth process: by finding the best split feature variable and the corresponding best split value, each node is iteratively split into two child nodes until the minimum node size is reached.

Fig.1 A simple illustration of CART (top) and RF (bottom)

A more detailed RF construction and prediction description is presented in Algorithm 1 below. In short, as mentioned earlier, the RF construction process is to build a large number of CARTs to fit the new training data sets resampled from the original training set through bagging. When constructing each CART, at each decision node, RF does not search for the best split variable among all feature variables but restricts the

candidates to m randomly selected features. This is the random node optimization mentioned above as a technique to reduce CART variance. After that, the best split feature among the m candidate features and corresponding best split value will be determined according to whether the labeled prediction results in the training data set can be most evenly split. Then the node will be divided into two sub-nodes, as will the re-sampled training data set. These steps will be repeated recursively for each decision node of a CART until the minimum node size n_{min} is reached. For the simple example in Fig. 1, the value of n_{min} and the number of CARTs in RF are set to be 2 and 10, respectively. Finally, at the prediction stage, the RF model makes prediction by output the average of the predictions generated by all the CARTs in it.

Algorithm 1. Growing and prediction stages of the RF model

Growing Stage:

1. Inputs of the RF model:
2. *a.* Training data set Z : N p -dimension samples with prediction result as label.
3. *b.* Parameter K : Number of CARTs in the RF model.
4. *c.* Parameter m : Number of split feature variables at each decision node.
5. *d.* Parameter n_{min} : Minimum decision node size.
6. For $k = 1$ to K :
7. *a.* Randomly draw a bootstrap training data set Z^* of size N^* from the original training data set Z (bagging).
8. *b.* Fit a CART T_k to the bootstrap training data set Z^* , by recursively repeating the steps *i* through *iii* below for each decision node of the CART until the minimum decision node size n_{min} is reached.
9. *i.* Randomly select m feature variables from the initial p feature variables.
10. *ii.* Determine the best split feature variable among the m feature variables and the best split value.
11. *iii.* Split the decision node into two sub-nodes.
12. Output the ensemble of CARTs $\{T_k\}_1^K$ as the RF model

Prediction Stage:

13. Let $\hat{Y}_k$ denotes the prediction result of the k th CART in RF model. Then the prediction result of the RF model can be calculated as the average of the prediction results of all the CARTs: $\hat{Y}_{RF} = \frac{1}{K} \sum_{k=1}^K \hat{Y}_k$
-

The merits of RF model are summarized as follows:

- 1) RF model is relatively easy to tune and optimize as it only requires 3 input parameters, including number of CARTs in RF model, number of split features for each node, and the minimum decision node size.
- 2) RF model can generate feature importance indices during its growing process, which turn out to be a nice estimate of variable relevancy.
- 3) With tree structures, RF model can expand itself by adding more branches to CARTs to adapt to new prediction environments that are different from the original training data

set. This evolved the RF online learning algorithm and made RF a good adaptive machine learning model for prediction.

4) The probability distribution of RF prediction results can provide input for MES optimal scheduling that considers multiple possible scenarios in the future. Other machine learning models, such as NN, cannot provide the probability distribution of prediction results.

The last two merits will be described in more detail in Section 3.

3 Probability distribution of prediction results and self-adaptability function of RF

This section describes in detail two important merits of the proposed adaptive RF model for prediction of MES demands and solar power, the probability distribution of prediction results and the online self-adaptability function.

3.1 Probability distribution of prediction results

Due to certain errors in the prediction process, the prediction will not be completely accurate. Users of the prediction model need to know how reliable the predicted results are. MES operators also need to consider a variety of scenarios that may occur in the future and comprehensively take them into consideration in the subsequent optimization instead of only considering the average future situation. Therefore, there is a need for a prediction model that can give the probability distribution of the predicted results and take them into account when scheduling MES operations. In this case, the operator can then consider the trade-offs in optimizing different possible future scenarios. The RF model can provide this useful function along with the prediction results.

In the RF prediction stage, the model output result is calculated as the average prediction of all the CARTs in it. Consider a CART T in the RF model generates predictions based on a set of observations Y . Denote the posterior distribution of the CART parameters θ by $\theta|Y$. Then the parameters $\hat{\theta}$ can be estimated based on a bootstrap set sampled from the observations Y , which can also approximate the posterior distribution of the CART parameters $\theta|Y$ according to [33]. Considering this, the CARTs in the RF model can approximate posterior distribution of the CART parameters $\theta|Y$. Thus, the predictions of the CARTs, $\hat{Y}_k$ (for $k = 1 \dots K$), reflect the posterior distribution of the prediction of the RF model and can be used to build the confidence interval attached to the prediction result.

Specifically, for building the confidence interval, the first step is to estimate the density and distribution of the prediction based on kernel density method as in Eq. (4) and then with the predefined confidence level α and the probability distribution in Eq. (5) the confidence interval attached to the prediction result can be generated.

$$f_Y(u) = \frac{1}{K\sigma\sqrt{2\pi}} \sum_{k=1}^K e^{-\frac{1}{2}(u-\hat{Y}_k)^2/\sigma^2} \quad (1)$$

$$F_Y(u) = \int_{-\infty}^u f_Y(v) dv \quad (2)$$

3.2 Online self-adaptability function

In addition to the confidence interval along with the prediction results, because the prediction environment encountered in the actual operation cannot be completely covered in the training data set, the prediction model needs to have an adaptive function to improve the prediction accuracy. Specifically, the RF model needs to be updated with new observations so that it can adapt to the latest prediction conditions. An online self-adaptability function is introduced to in this work to help update the RF model with new observations, as shown in Algorithm 2 below. The main ideas of the online self-adaptability function are summarized as follows:

For each CART in the adaptive RF model, the new observation, which are the values of the new collected feature variables, will be processed λ times repeatedly. This practice aims to simulate the bagging sampling procedure in the RF learning algorithm. The updating process is as follows. First, the new data is passed down from the first decision node to one of the leaf nodes (prediction results) to which the observation belongs. Then, based on two criterions the adaptive RF model decides whether to split the leaf node of a CART into two child nodes: 1) whether some leaf nodes have received enough new data that exceeds a predefined threshold, and 2) if a splitting is made for a leaf node, whether possible improvement in prediction accuracy can be achieved.

Algorithm 2. Online self-adaptability function of the adaptive RF model

-
-
1. Online self-adaptability function inputs:
 2. *a.* New observed data Z_R : N_R p -dimension samples with label as prediction results.
 3. *b.* Parameter K : Number of CARTs in RF model
 4. *c.* Parameter σ : Threshold of minimum number of samples received at leaf node.
 5. *d.* Parameter ζ : Prediction accuracy improvement in percentage.
 6. // For all CARTs in RF model
 7. **for** k from 1 to K **do**
 8. **for** i from 1 to λ **do**
 9. $j = \text{findLeaf}(N_R)$.
 10. **if** $|R_j| > \sigma$ and $\exists s \in S: \Delta\gamma_k(R_j, s) > \zeta$ **then**
 11. Find the best split value at leaf node j :
 $s_j = \text{argmax}_{s \in S} \Delta\gamma_k(R_j, s)$
 12. CreateLeftChildNode.
 13. CreateRightChildNode.
 14. **end if**
 15. **end for**
 16. **end for**
 17. **Output** the updated RF.
-
-

4 Simulation results and analysis

In the simulation section, we will use the data of Stone Edge Farm, which is one of the world's first-class MES test beds located in Sonoma, California, to show the merits of the proposed adaptive RF model in predicting of demands and solar

power. Fig.2 below shows the aerial view of Stone Edge Farm MES. MATLAB R2021a is employed in an Intel Processor 5Y70 CPU 1.3 GHz environment to handle all the simulations in this section.

Fig.2 Aerial view of Stone Edge Farm MES.

For the proposed adaptive RF model, the selected features for the prediction of each kind of demand (electric, heating, cooling, and hydrogen) of Stone Edge Farm MES after certain hours include:

- Indicator of the day (one of Monday through Sunday)
- Current and predicted temperature after certain hours
- Demands at present, one hour ago, three hours ago, six hours ago, one day ago, one week ago, and one month ago

Similarly, the selected features for the prediction of solar power after certain hours include:

- Weather information after certain hours (sunny, cloudy, showers, etc.)
- Current and predicted temperature after certain hours
- Solar power at present, one hour ago, three hours ago, six hours ago, one day ago, one week ago, and one month ago

These features were chosen because that they are found to be the most influential features for the corresponding predictions according to the literatures. The historical data of Stone Edge Farm MES from September 1st to November 15th, 2019, is used as the training data set, and the data from November 16th to November 30th, 2019, is used as the test data set. Table 1 below shows the selected parameters of the proposed adaptive RF model for the simulations (different parameter combinations have been compared to choose the optimal parameter values).

As mentioned earlier, a key feature provided by the RF model is the distribution of prediction results, which can be used to form the corresponding confidence interval. Users can use this feature to evaluate the reliability of prediction results, and it can also be used to generate a variety of possible scenarios that may occur in the future for MES scheduling optimization. Fig. 3 to Fig. 7 below demonstrated the predicted hourly demands and solar power for the next 24 hours of Stone Edge Farm MES at 8 am on December 17, 2019, based on the proposed adaptive RF model. The confidence interval of the

prediction results is set to be 90% in this illustration, which means 90% of the prediction results of all the CARTs in the adaptive RF model are selected and are shown in lighter color in the same figures.

Table 1 Adaptive RF model parameters

Parameter	Value
K	500
m	3
n_{min}	5
λ	20
α	90%
δ	10
ξ	0.1

Fig.5 Predicted cooling power demand for the next 24 hours

Fig.3 Predicted electric power demand for the next 24 hours

Fig.6 Predicted hydrogen power demand for the next 24 hour

Fig.4 Predicted heating power demand for the next 24 hours

Fig.7 Predicted solar power for the next 24 hours

To verify the effectiveness of the proposed adaptive RF model in improving the prediction accuracy, in the second part of the simulation the proposed model is compared with traditional RF model without the online self-adaptability function, NN model, and ARMA model, which are typical models for demands and solar power predictions according to the literatures. The four prediction models are denoted as models A, B, C and D throughout the remaining part of the simulation.

The selected features and training data set of models B and C would be the same as model A. Since the ARMA model adopts the recent data to linearly predict the future value of the same variable, previous 30 days demands and solar power historical data of Stone Edge Farm MES are adopted in this simulation to train the model D. The selected parameters of benchmark prediction models B, C and D can be found in Table 2 below. For model B, same RF parameter values are adopted as model A except the parameters about online self-adaptability are removed. For model C, since the NN model has 10 input features and 1 output predicted value (demand or solar energy certain hours later), the input layer will have 10 neurons and the output layer will have 1 single output neuron. According to our experiments, the optimal NN structure is optimized to three hidden layers, where the first hidden layer has 8 neurons, the second layer has 6 neurons, and the third layer has 4 neurons. For model D, the orders of autoregressive part and moving average part are optimized to 145 and 30, respectively, to generate the highest prediction accuracy.

As in Table 3 below, the prediction models were tested and compared with the actual data of Stone Edge Farm MES in December, and the average absolute percentage errors (MAPE) for the predictions of future demands and solar power for 1 hour and 24 hours later are shown. It can be seen from the results that the proposed adaptive RF model is better than other benchmark prediction models in terms of prediction accuracy. This is because whenever a new observation is detected, the proposed adaptive RF model can be adjusted to the latest prediction conditions. Therefore, the prediction accuracy can be improved. However, the benchmark prediction models do not have this capability.

5 Conclusion

This paper proposes an adaptive RF model to predict the demands and solar power of a real MES, Stone Edge Farm, in California. The proposed adaptive RF model can provide a probability distribution of the prediction results. This allows users to consider a variety of scenarios that may occur in the future for further system operation optimization and help users evaluate the reliability of the results. Besides, an online self-adaptability function is implemented with the model so it can adapt to the new forecasting environment when new observations are detected. To the authors' knowledge, this is the first detailed paper describing RF applications in MES optimal scheduling problem. Besides reliability visualization through confidence interval, the simulations show that the adaptive RF is better than the benchmark prediction models in terms of prediction accuracy.

Table 2 Selected parameters of benchmark prediction models B, C and D

Parameter	Value
K	500
m	3
n_{min}	5
N_I	10
N_O	1
N_{HL}	3
N_{L1}	8
N_{L2}	6
N_{L3}	4
μ	145
β	30

Table 3 Comparison of different forecasting models

	MAPE of 1 hour	MAPE of 24 hours
Model A	0.526%	4.356%
Model B	0.535%	4.812%
Model C	0.539%	4.843%
Model D	0.817%	6.724%

6 References

- [1] Li, Y., Zou, Y., et al.: 'Optimal stochastic operation of integrated low-carbon electric power, natural gas and heat delivery system', *IEEE Trans. Sustain. Energy*, 2017, **9**, (1), pp. 273–283
- [2] Feng, Z., Tang, W., et al.: 'Bi-level allocation of carbon emission permits based on clustering analysis and weighted voting: A case study in China', *Appl. Energy*, 2018, **228**, pp. 1122–1135
- [3] Pan, Z., Guo, Q., Sun, H.: 'Feasible region method based integrated heat and electricity dispatch considering building thermal inertia', *Appl. Energy*, 2017, **192**, pp. 395–407
- [4] Geidl, M., Andersson, G.: 'Optimal power flow of multiple energy carriers', *IEEE Trans. Power Syst.*, 2007, **22**, (1), pp. 145–155
- [5] Mei, J., Wang, X., Kirtley, J.L.: 'Optimal scheduling of real multi-carrier energy storage system with hydrogen-based vehicle applications', *IET Renewable Power Generation*, 2020, **14**, (3), pp. 381–388
- [6] Mei, J., Zuo, Y., Lee, C., et al.: 'Stochastic optimization of multi-energy system operation considering hydrogen-based vehicle applications', *Advan. Appl. Energy*, 2021, **2**, 100031
- [7] Wang, Y., Zhang, N., et al.: 'Mixed-integer linear programming-based optimal configuration planning for energy hub: Starting from scratch', *Appl. Energy*, 2018, **210**, pp. 1141–1150
- [8] Moser, A., Muschick, D., et al.: 'A MILP-based modular energy management system for urban multi-energy systems: Performance and sensitivity analysis.', *Appl. Energy*, 2020, **261**, pp. 114342
- [9] Kampouropoulos, K., Andrade, F.: 'Multiobjective optimization of multi-carrier energy system using a combination of ANFIS and genetic algorithms', *IEEE Trans. Smart Grid*, 2016, **9**, (3), pp. 2276–2283
- [10] Mayer, M.J., Szilágyi, A., et al.: 'Environmental and economic multi-objective optimization of a household level hybrid renewable energy system by genetic algorithm', *Appl. Energy*, 2020, **269**, pp. 115058
- [11] Wu, L., Shahidehpour, M.: 'Optimal coordination of stochastic hydro and natural gas supplies in midterm operation of power systems', *IET Gener. Transm. Distrib.*, 2011, **5**, (5), pp. 577–587

- [12] Li, X., Zhang, R., Bai, L., et al.: 'Stochastic low-carbon scheduling with carbon capture power plants and coupon-based demand response', *Appl. Energy*, 2018, **210**, pp. 1219–1228
- [13] Qu, G., Mei, J., He, D.: 'Short-term wind power forecasting based on numerical weather prediction adjustment', *IEEE International Conference on Industrial Informatics (INDIN)*, 2013, Jul 29, pp. 453–457
- [14] Kong, W., Dong, Z., Jia, Y., et al.: 'Short-term residential load forecasting based on LSTM recurrent neural network', *IEEE Trans. Smart Grid*, 2017, 10, (1), pp. 841–51
- [15] Li, Y., Wang, R., Yang, Z.: 'Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting', *IEEE Trans. Sustain. Energy*, 2021.
- [16] Mansouri, M., Fezai, R., Trabelsi, M., et al.: 'Reduced Gaussian process regression based random forest approach for fault diagnosis of wind energy conversion systems', *IET Renewable Power Generation*, 2021, **15**, (15), pp.3612–3621
- [17] Farooq, F., Nasir, A.M., Khan, K., Rehan, S.M., Faisal, J.M., Aslam, F., Alyousef, R.: 'A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC)', *Applied Sciences*, 2020, **10**, (20), 7330
- [18] Petkovic, D., Barlaskar, S.H., Yang, J., Todtenhoefer, R.: 'From explaining how random forest classifier predicts learning of software engineering teamwork to guidance for educators', *2018 IEEE Frontiers in Education Conference (FIE)*, 2018, pp. 1–7
- [19] Ahmed, A.M., Deo, R.C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., Yang, L.: 'Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity', *Journal of Hydrology*, 2021, **599**, 126350
- [20] Yoo, C., Han, D., Im, J., Bechtel, B.: 'Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images', *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, **157**, pp. 155–170
- [21] Rather, T.A., Kumar, S., Khan, J.A.: 'Multi-scale habitat selection and impacts of climate change on the distribution of four sympatric meso-carnivores using random forest algorithm', *Ecological Processes*, 2020, **9**, (1), pp. 1–7
- [22] Zhang, L., Alharbe, N.R., Luo, G., Yao, Z., Li, Y.: 'A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction', *Tsinghua Science and Technology*, 2018, **23**, (4), pp. 479–492
- [23] Li, L., Prato, C.G., Wang, Y.: 'Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: A sequential approach of multivariate imputation by chained equations and random forest classifier', *Accident Analysis & Prevention*, 2020, **146**, 105744
- [24] Li, L., Prato, C.G., Wang, Y.: 'Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: A sequential approach of multivariate imputation by chained equations and random forest classifier', *Accident Analysis & Prevention*, 2020, **146**, 105744
- [25] Thonfeld, F., Steinbach, S., Muro, J., Kirimi, F.: 'Long-term land use/land cover change assessment of the Kilombero catchment in Tanzania using random forest classification and robust change vector analysis', *Remote sensing*, 2020, **12**, (7), 1057
- [26] Mei, J., He, D., Harley, R., et al.: 'A random forest method for real-time price forecasting in New York electricity market', *2014 IEEE PES General Meeting/Conference & Exposition*, 2014, Jul 27, pp. 1–5
- [27] Hümberlin, O., Hobi, L., Fluder, R.: 'Rich Cities, Poor Countryside? Social Structure of the Poor and Poverty Risks in Urban and Rural Places in an Affluent Country. An Administrative Data based Analysis using Random Forest', *University of Bern, Department of Social Sciences*, 2021
- [28] Gounaridis, D., Chorianopoulos, I., Symeonakis, E., Koukoulas, S.: 'A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales', *Science of the Total Environment*, 2019, **646**, pp. 320–335
- [29] Guo, C., Zhang, J., Liu, Y., Xie, Y., Han, Z., Yu, J.: 'Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform', *IEEE Access*, 2020, **8**, pp. 59247–59256
- [30] Wang, J., Shi, L.: 'Prediction of medical expenditures of diagnosed diabetics and the assessment of its related factors using a random forest model, MEPS 2000–2015', *International Journal for Quality in Health Care*, 2020, **32**, (2), pp. 99–112
- [31] Geetha, R., Sivasubramanian, S., Kaliappan, M., Vimal, S., Annamalai, S.: 'Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier', *Journal of medical systems*, 2019, **43**, (9), pp. 1–9
- [32] Mei, J., He, D., Harley, R.G., Habetler, T.G.: 'Random forest based adaptive non-intrusive load identification', *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 1978–1983
- [33] Hastie, T., Tibshirani, R., Friedman, J.: 'The elements of statistical learning: prediction. inference and data mining', 2009