IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Smart City Energy Prediction Using Random Forest

[1]Mr. Pritam Ahire, [2]Sejal Katkar, [3]Riddhi Barhate, [4]Samiksha Asodekar

[1]Professor, [2]Student, [3]Student, [4]Student

[1]Department Computer Engineering,

[1]Nutan Maharashtra Institute of Engineering and Technology, Pune, India

Abstract: Accurate energy demand forecasting is crucial for optimizing power distribution, reducing energy waste, and improving resource efficiency. Traditional forecasting methods often suffer from limitations in accuracy and scalability. Using a variety of algorithms, such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), k-Nearest Neighbours (kNN), and XGBoost, this study suggests a machine learning-based method to forecast energy consumption. Real-world data from commercial buildings in Malaysia is utilized, considering the influence of environmental factors such as temperature. The dataset is pre-processed to deal with missing values and outliers in order to guarantee dependable model training. The predictive performance of different models is evaluated using metrics such as RMSE, NRMSE, MAE, and MAPE. The results indicate that machine learning models provide significant improvements in forecast accuracy, with ANN and KNN outperforming traditional approaches. This study contributes to energy management by offering an interpretable and scalable solution for demand prediction in smart grids.

***Index Terms* - Energy Demand Forecasting, Machine Learning, Artificial Neural Networks, k-Nearest Neighbors, Support Vector Machine, Predictive Modeling, RMSE.**

I. INTRODUCTION

Forecasting energy demand is essential to effective energy management, especially when it comes to smart buildings and urban energy usage.

The emergence of smart buildings has leveraged advancements in computational and communication architectures to enhance energy efficiency through automated control and intelligent forecasting systems [1]. These smart systems not only optimize energy usage but also contribute to economic savings and sustainability efforts by minimizing energy wastage [2].

Nearly 40% of all primary energy use in developed countries comes from buildings, making them a major source of energy consumption [3].

In countries like Malaysia, energy consumption has been rising steadily due to population growth and increased industrialization, leading to a projected demand of 116 million tons of oil equivalents (mtoe) in the near future [4]. Such trends necessitate advanced forecasting methodologies to balance energy supply and demand effectively.

Energy demand forecasting has made extensive use of conventional statistical techniques, such as time series analysis. Nevertheless, these methods frequently encounter constraints when handling non-stationary time series, high-dimensional features, and sizable datasets [1].

The incorporation By identifying intricate dependencies in energy consumption data, machine learning (ML) techniques have become a viable substitute that offer better predictive performance. Recurrent neural networks (RNNs) and autoregressive integrated moving averages (ARIMA) are two models that have proven to be more accurate than traditional techniques at predicting electricity demand [2].

Furthermore, patterns of energy consumption are significantly impacted by climate change. Increased demand for cooling systems due to rising global temperatures causes variations in electricity consumption. Global average surface temperatures have risen by about 0.85°C since 1880, according to the Intergovernmental Panel on Climate Change (IPCC), and more increases are predicted by the end of the century [3]. Energy demand is impacted by these climate variations, so accurate forecasts require strong forecasting models that take operational, socioeconomic, and environmental factors into account [4].

The goal of this project is to use cutting-edge machine learning techniques in a cloud-based machine learning platform to create a predictive model for energy consumption in smart commercial buildings. Enhancing forecast accuracy and tackling issues like missing data and computational efficiency are the main goals of the study. This work aims to give policymakers, energy providers, and industrial stakeholders insights to improve energy planning and sustainability strategies by utilising cloud computing and machine learning techniques.

II. LITERATURE SURVEY

2.1 Machine Learning Prediction Methodology

The ability of machine learning to model intricate relationships between past data and future power consumption has led to its widespread adoption in energy demand forecasting. Based on historical power consumption data, predictive models are created and system parameters are estimated using a data-driven methodology [5]. Several machine learning methods have been shown to be successful in predicting energy demand in earlier research. For example, using hourly electricity load inputs and weather forecasts, Support Vector Machines (SVM) have been used to forecast building-level electricity consumption, with an RMSE of 15.2% and a Mean Bias Error (MBE) of 7.7% [6]. Similar to this, the k-Nearest Neighbour (k-NN) model has been used to forecast power demand in smart buildings. It classifies similar cases and makes short-term predictions by utilising historical load curves [7].

For short-term load forecasting, additional machine learning techniques have also been used, including the Gaussian Process, REPTree, Radial Basis Function (RBF) regressor, and Multi-Layer Perceptron (MLP). With a Mean Absolute Percentage Error (MAPE) of 0.96%, MLP has proven to be the most accurate of these [8]. Regression-based machine learning techniques have also been investigated, such as k-NN, Random Forest, Decision Tree, Linear Regression, Support Vector Regression, and Random Forest. With an overall accuracy of 85.7%, studies show that LR and SVR models produce the most accurate predictions [7].

2.2 Management of Missing Data

Handling missing data is a crucial aspect of energy demand forecasting. Traditional methods, such as deleting incomplete data, often lead to biased estimations and reduced model performance [6]. Instead, imputation methods, which estimate missing values based on observed data, are preferred. Several techniques have been proposed, including Mean Value Imputation, Last Observation Carried Forward, Maximum Likelihood Estimate (MLE), and Multiple Imputation (MI) [5].

Numerous studies have compared advanced imputation techniques like Probabilistic Principal Component Analysis (PPCA) and Multiple Imputation Using Chained Equations (MICE). In large datasets, studies have demonstrated that PPCA performs better than MICE, successfully imputing 65% of missing variables as opposed to 38% for MICE [8]. These results emphasise how crucial it is to choose the best imputation methods depending on the features of the dataset and the needs of the application.

2.3 Employment of Cloud-Based Prediction Modeling

With the rise of big data and real-time analytics, cloud-based machine learning platforms have become essential for scalable and efficient energy demand forecasting. Platforms such as Apache Spark's Machine Learning Library (Spark MLlib), TensorFlow, and Microsoft Azure Machine Learning Studio (Azure ML) provide robust environments for handling large-scale energy consumption data [6].

Because of its ability to detect consumption trends with little computational resources, Apache Spark MLlib has been used extensively for analysing patterns of energy consumption in sizable time-series datasets. [7]

TensorFlow, primarily focused on deep learning and reinforcement learning, has been employed for complex energy prediction tasks, such as Convolutional Neural Networks (CNN)-based modeling for energy demand forecasting [8]. Additionally, Microsoft Azure ML is widely used by enterprises due to its cloud-based predictive analytics capabilities, requiring minimal hardware investments while supporting multiple programming languages, including Python and R [5].

III. PROBLEM DEFINITION

Dynamic Energy Landscape: The energy landscape is extremely dynamic due to the quick integration of renewable energy sources, technological breakthroughs, and changes in consumer behaviour. Conventional forecasting techniques might find it difficult to adjust to the complex and changing patterns brought about by these developments. Complex Relationships and Nonlinearities: Nonlinearities and complex relationships in data on electricity consumption are frequently missed by traditional time-series models. The need for forecasting models that can manage these dynamic and nonlinear patterns is increasing as the energy mix diversifies and becomes more complex. Influence of External Factors: The demand for electricity is greatly influenced by external factors, including weather, economic indicators, and emerging technologies. Current forecasting approaches may not fully exploit the potential of incorporating these factors, leading to suboptimal predictions.

IV. Proposed Methodology

This research aims to develop an energy demand forecasting model using machine learning techniques. The methodology consists of four major steps: normality testing, data pre-processing, model development, and model evaluation.

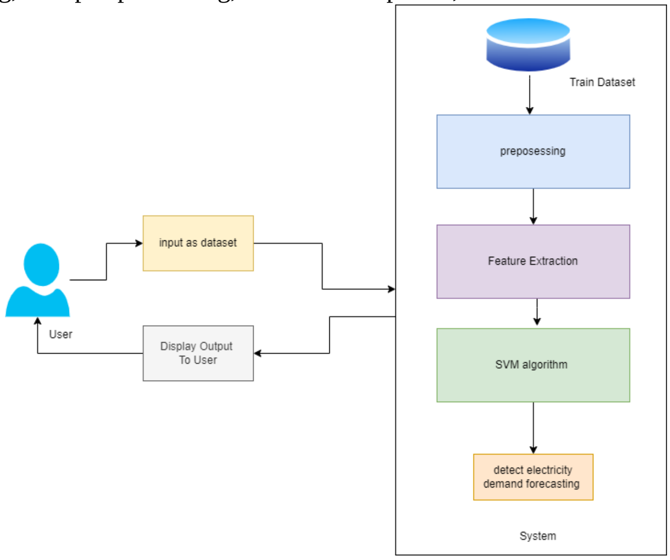

Fig 1 : System Architecture

Module 1 : Data Collection and Pre-processing
Module 2 : Model Development
Module 3 : Model Evaluation

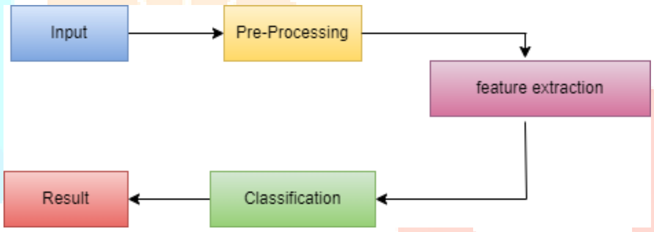

Fig 2 : Flowchart

1. Data Collection and Pre-processing

An IoT-enabled commercial building in Malaysia's Klang Valley that measures energy consumption per minute provided the dataset for this study. The data can be extracted in CSV format from an open-source web server after being mapped in accordance with Tenaga National Berhad's (TNB) specifications [9]. The dataset's target variable is energy demand, and its features include power factor, voltage, and current.

The raw data is preprocessed before model training, and this includes:

- **Handling missing values:** Missing data is imputed using statistical methods.
- **Feature selection:** Only relevant variables (power factor, voltage, and current) are retained.
- **Normalization:** Features are scaled to improve model performance [10].

2. Model Development

Three machine learning algorithms—Random Forest, K-Nearest Neighbors (KNN), and Support Vector Regression (SVR)—are used for energy demand prediction. These models have been chosen due to their effectiveness in previous studies on energy consumption forecasting [11].

- **Random Forest (RF):** To increase prediction accuracy, RF is an ensemble learning technique that combines several decision trees with Bootstrap Aggregation (Bagging). It improves generalisation and lessens overfitting [10].

- **K-Nearest Neighbors(KNN):** KNN regression is a non-parametric technique that uses the average of the K-nearest data points to estimate output. Performance is greatly impacted by the choice of K; a higher K may oversimplify the model, while a lower K may result in overfitting [11].

- **Support Vector Regression (SVR):** SVR finds an optimal hyperplane to minimize prediction errors while using kernel functions to capture complex relationships in the data [12].

3. Model Evaluation

The trained models are evaluated using multiple performance metrics:

- The standard deviation of prediction errors is measured by the Root Mean Squared Error (RMSE).
- The average absolute differences between the expected and actual values are assessed using the Mean Absolute Error (MAE).
- **R-squared (R^2):** Shows the percentage of the dependent variable's variance that the model can account for.

Hyperparameter tuning is performed using cross-validation to optimize model performance. The entire implementation is carried out using Python with libraries such as Scikit-Learn and XGBoost.

By following this structured approach, the study ensures a robust prediction framework that enhances energy demand forecasting accuracy and reliability.

V. Advantages

The adoption of Random Forest (RF) for energy demand forecasting provides several key benefits over traditional forecasting methods:

1. **Handling Nonlinear Relationships:** RF captures complex dependencies in energy consumption, which is crucial for environments with fluctuating energy demand patterns and renewable energy integration. Unlike linear models, RF can model non-linear relationships effectively [13].
2. **Robustness to Overfitting:** As an ensemble learning method, RF mitigates overfitting by averaging multiple decision trees, leading to better generalization across different datasets [14].
3. **Ability to Handle High-Dimensional Data:** RF processes datasets with multiple features, such as weather conditions, economic indicators, and demographic variables, without requiring extensive feature engineering [15].
4. **Interpretability and Feature Importance:** RF provides feature importance rankings, helping to identify key factors influencing energy demand and aiding in transparent decision-making [16].

VI. Existing System Disadvantages

Traditional forecasting models and existing machine learning approaches face limitations in handling modern energy systems.

1. **Inability to Handle Nonlinearity:** Linear models like ARIMA fail to capture complex dependencies, reducing accuracy, especially with renewable energy sources that exhibit nonlinear patterns [17].
2. **Poor Adaptability:** Traditional models assume stable historical patterns, making them ineffective for evolving energy markets influenced by consumer behavior and policy changes [18].
3. **Limited Handling of High-Dimensional Data:** These models struggle with large datasets, often oversimplifying relationships and ignoring critical external factors like economic and weather data [19].
4. **Lack of Flexibility and Robustness:** Many existing methods are not adaptable to noisy or missing data and require extensive manual tuning, making them less efficient than modern techniques like Random Forest [20].

VII. Application

Forecasting energy demand is essential for efficient energy use and sustainability in a variety of fields, including grid management and policymaking. Forecasting accuracy has increased dramatically with the use of machine learning techniques, especially Random Forest (RF), which makes it a useful tool in contemporary energy systems.

1. **Grid Management & Load Forecasting:** Maintaining grid stability and achieving the best possible supply and demand balance depend on accurate load forecasting. To produce accurate short- and long-term forecasts, RF-based models examine past energy usage, weather trends, and economic indicators. With the increasing integration of renewable energy sources, this aids grid operators in efficiently allocating resources, reducing power outages, and optimising energy distribution.
2. **Renewable Energy Integration:** Because renewable energy sources like solar and wind are weather-dependent, their growing use adds uncertainty to energy systems. RF models help predict fluctuations in renewable energy generation and consumption by capturing complex dependencies between meteorological conditions and energy demand. This enhances the coordination between renewable and conventional power generation, improving overall grid efficiency and stability.
3. **Energy Efficiency Optimization:** Forecasting energy demand allows industries, businesses, and households to optimize energy consumption patterns and reduce inefficiencies. Better energy conservation results from the use of RF-based models, which pinpoint periods of peak demand and recommend load management techniques. In smart grids and intelligent buildings, RF enables real-time energy monitoring, facilitating automated adjustments in heating, cooling, and lighting systems to enhance efficiency and reduce operational costs.
4. **Policy Planning & Decision Support:** Governments and energy regulators rely on accurate demand forecasts to formulate energy policies, infrastructure development plans, and pricing strategies. RF-driven forecasting models incorporate multiple socio-economic and environmental factors, enabling policymakers to anticipate future energy needs and implement data-driven strategies for sustainable energy management. This is particularly valuable for urban planning, renewable energy investments, and demand-side management programs.

VIII. Conclusion

This study developed an energy demand forecasting model leveraging machine learning techniques to improve prediction accuracy. The methodology included data pre-processing, feature selection, and model development using Random Forest, KNN, and SVR. The results indicated that ML-based models, particularly SVR, performed effectively in energy demand forecasting, aligning with findings from previous research [20]. Furthermore, incorporating auxiliary factors like temperature demonstrated potential for enhancing predictive performance, as highlighted in existing studies [21]. The cloud-based implementation ensured scalability and robustness, reinforcing its viability for smart building energy management [22]. Future research can explore ensemble methods and additional external variables, such as economic indicators, to enhance model accuracy [23].

IX. Future Scope

Future research in energy demand forecasting can focus on integrating hybrid machine learning models, real-time data processing, and uncertainty quantification to enhance predictive accuracy. Combining Random Forest with deep learning or support vector machines can improve forecasting performance. Additionally, leveraging IoT data and big data analytics can enable more precise demand predictions. Real-time decision support systems for smart grids and energy storage integration will further optimize energy management. Probabilistic forecasting methods and risk management strategies can also improve reliability in handling demand fluctuations, ensuring a more efficient and sustainable energy system.

References

[1] Ahire, Pritam Ramesh, and K. Ulaga Priya. "Monitoring Body Mass Index (BMI) Pre & Post Covid-19 Outbreak: A Comprehensive study in Healthcare." 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon). IEEE, 2024.

[2] Ahire, Pritam. "Predictive and Descriptive Analysis for Healthcare Data." A Handbook on Intelligent Health Care Analytics - Knowledge Engineering with Big Data, Scrivener Publishing, 2021. Available at: https://www.wiley.com/enus/Handbook+on+Intelligent+Healthcare+Analytics%3A+Knowledge+Engineering+with+Big+Data-p-9781119792536.

[3] Ahire, Pritam, et al. "LSTM based stock price prediction." International Journal of Creative Research Thoughts, vol. 9, no. 2, 2021, pp. 5118-5122.

[4] Ahire, Pritam R., and Preeti Mulay. "Discover compatibility: Machine learning way." Journal of Theoretical & Applied Information Technology, vol. 86, no. 3, 2016.

[5] Ahire, Pritam R., Rohini Hanchate, and Vijayakumar Varadarajan. "Indigenous Knowledge in Smart Agriculture." Advanced Technologies for Smart Agriculture, River Publishers, 2024, pp. 241-258.

[6] Hanchate, R., & Anandan, R. "Medical Image Encryption Using Hybrid Adaptive Elliptic Curve Cryptography and Logistic Map-based DNA Sequence in IoT Environment." IETE Journal of Research, 2023, pp. 1–16. Available at: https://doi.org/10.1080/03772063.2023.2268578.

[7] Ahire, Pritam Ramesh, Rohini Hanchate, and K. Kalaiselvi. "Optimized Data Retrieval and Data Storage for Healthcare Applications." Predictive Data Modelling for Biomedical Data and Imaging, River Publishers, pp. 107-126.

[8] Sajiharan, Sabaretnam & Singh, Kishan. (2023). LEVERAGING IOT AND MACHINE LEARNING FOR ENERGY-EFFICIENT SMART BUILDINGS: A COMPREHENSIVE ANALYSIS. European Chemical Bulletin. 12. 2111-2117. 10.31838/ecb/2023.12.s3.264.

[9] Seyed Azad Nabavi, Sahar Mohammadi, Naser Hossein Motlagh, Sasu Tarkoma, Philipp Geyer, Deep learning modeling in electricity load forecasting: Improved accuracy by combining DWT and LSTM, Energy Reports, Volume 12, 2024, Pages 2873-2900, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2024.08.070.

[10] Yixuan Wei, Xingxing Zhang, Yong Shi, Liang Xia, Song Pan, Jinshun Wu, Mengjie Han, Xiaoyun Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews, Volume 82, Part 1, 2018, Pages 1027-1047, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2017.09.108.

[11] Chandramowli, Shankar & Felder, Frank. (2013). Impact of Climate Change on Electricity Systems and Markets - A Review of Models and Forecasts. Sustainable Energy Technologies and Assessments. 5. 10.2139/ssrn.2251167.

[12] Tardioli, Giovanni & Kerrigan, Ruth & Oates, Mike & O Donnell, James & Finn, Donal. (2015). Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level. Energy Procedia. 78. 3378-3383. 10.1016/j.egypro.2015.11.754.

[13] Bilal Abu-Salih, et al., Short-term renewable energy consumption and generation forecasting: A case study of Western Australia, Heliyon, Volume 8, Issue 3, 2022, e09152, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2022.e09152.

[14] Calvillo, Christian & Sánchez-Miralles, A. & Villar, Jose. (2016). Energy management and planning in smart cities. Renewable and Sustainable Energy Reviews. 55. 273-287. 10.1016/j.rser.2015.10.133.

[15] Kaur, Devinder & Islam, Shama & Mahmud, Md. Apel & Dong, Z.Y.. (2020). Energy Forecasting in Smart Grid Systems: A Review of the State-of-the-art Techniques. 10.48550/arXiv.2011.12598.

[16] Ryu, Seungjun & Egnal, Andre & Karagiorgis, George & Charalambous, Costas & Strbac, Goran & Chrysochos, Alexandros. (2020). A data-driven approach for probabilistic energy demand forecasting using the quantile regression forest method. Energies. 13. 3230. 10.3390/en13123230.

[17] Amin-Naseri, Mohammad & Hejazi, Seyed & Moloodi, Amir. (2015). Forecasting electricity consumption using multi-layer perceptron and improved particle swarm optimization. Journal of Artificial Intelligence and Soft Computing Research. 5. 197-209. 10.1515/jaiscr-2015-0027.

[18] Rajasekaran, J. & Arunkumar, K. & Vinoth Kumar, M. & Sanjeevi, J.. (2021). A deep learning-based approach for short-term power consumption forecasting in smart cities. Journal of Green Engineering. 11. 11393-11412.

[19] Chen, Yu & Wang, Lei & Shen, Junjie & Zhang, Haoran & Zhou, Ping & Li, Wei & Wang, Chuan & Zhang, Yifan. (2021). A Hybrid Model Combining CNN and BiLSTM for Short-Term Electric Load Forecasting. IEEE Access. 9. 32132-32144. 10.1109/ACCESS.2021.3060063.

[20] Hamid Reza Shaker, Filippo Bianchini, Jacopo Torriti, Roberto Poli, "Forecasting residential electricity demand: A review of methods and applications," Renewable and Sustainable Energy Reviews, Volume 75, 2017, Pages 287-301, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2016.10.078.

[21] Tso, Geoffrey & Yau, Kelvin. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree, and neural networks. Energy. 32. 1761-1768. 10.1016/j.energy.2006.11.010.

[22] Panapakidis, Ioannis & Dagoumas, Athanasios. (2016). Day-ahead electricity price forecasting via the application of artificial neural network based models. Applied Energy. 172. 132-151. 10.1016/j.apenergy.2016.03.089.

[23] Kavousian, Amir & Rajagopal, Ram & Fischer, Martin. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy. 55. 184-194. 10.1016/j.energy.2013.03.086.

[24] Hongyu Zhang, Bo Chen, Ying Li, Junwei Geng, Cong Li, Wenyi Zhao, Haobo Yan, Research on medium- and long-term electricity demand forecasting under climate change, Energy Reports, Volume 8, Supplement 4, 2022, Pages 1585-1600, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.02.210.

[25] Ammar Kamoona, Hui Song, Kian Keshavarzian, Kedem Levy, Mahdi Jalili, Richardt Wilkinson, Xinghuo Yu, Brendan McGrath, Lasantha Meegahapola, Machine learning based energy demand prediction, Energy Reports, Volume 9, Supplement 12, 2023, Pages 171-176, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2023.09.151.

[26] Prajowal Manandhar, Hasan Rafiq, Edwin Rodriguez-Ubinas, Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods, Energy Reports, Volume 9, 2023, Pages 2757-2776, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2023.01.094.

[27] Ali Nikseresht, Hamidreza Amindavar, Energy demand forecasting using adaptive ARFIMA based on a novel dynamic structural break detection framework, Applied Energy, Volume 353, Part A, 2024, 122069, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2023.122069.

[28] Mahziyar Dostmohammadi, Mona Zamani Pedram, Siamak Hoseinzadeh, Davide Astiaso Garcia, A GA-stacking ensemble approach for forecasting energy consumption in a smart household: A comparative study of ensemble methods, Journal of Environmental Management, Volume 364, 2024, 121264, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2024.121264.

[29] Paudel, Subodh & Nguyen, Phuong & Kling, Wil & Elmitri, Mohamed & Lacarrière, Bruno & Corre, Olivier. (2015). Support Vector Machine in Prediction of Building Energy Demand Using Pseudo Dynamic Approach.

[30] Shataee, Shaban & Kalbi, Siavash & Fallah, Asghar & Pelz, Dieter. (2012). Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. International Journal of Remote Sensing - INT J REMOTE SENS. 33. 6254-6280. 10.1080/01431161.2012.682661.