REPUBLIC OF TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT

HYDROELECTRIC POWER FORECASTING VIA TREE-BASED MACHINE LEARNING ALGORITHMS

BEKTAŞ AYKUT ATALAY

M.Sc.

ADANA 2024

REPUBLIC OF TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT

HYDROELECTRIC POWER FORECASTING VIA TREE-BASED MACHINE LEARNING ALGORITHMS

BEKTAŞ AYKUT ATALAY

M.Sc.

THESIS ADVISOR
ASST. PROF. DR. KASIM ZOR

ADANA, 2024

DECLARATION OF CONFORMITY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

24/06/2024

[Signature]

Bektaş Aykut ATALAY

ÖZET

AĞAÇ TABANLI MAKİNE ÖĞRENMESİ ALGORİTMALARIYLA HİDROELEKTRİK GÜÇ TAHMİNİ

Bektaş Aykut ATALAY
Yüksek Lisans, Elektrik-Elektronik Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Kasım ZOR

Haziran 2024, 48 sayfa

Hidroelektrik enerji, en eski ve temel yenilenebilir enerji kaynağıdır. Türkiye'de hızlı ekonomik ve nüfus artışı nedeniyle enerji talebi sürekli artmaktadır ve bu durum hidroelektrik enerjiyi Türkiye'nin enerji kaynakları için kritik hale getirmektedir. Hidroelektrik santral planlaması ve uygulaması, devlet ve enerji işletmeleri için büyük önem taşımaktadır. Hidroelektrik enerji, mevsimsel bağımlılığı nedeniyle tahmin algoritmaları için uygundur. Hidroelektrik enerjinin doğru tahmin edilmesi, karbon emisyonlarının azaltılması, üretim verimliliği ve çevresel sürdürülebilirlik sağlar. Bu tez, kapasitesi 100 MW'ın üzerindeki faal bir hidroelektrik güç santralinin güç üretimini tahmin etmek için ağaç tabanlı makine öğrenmesi algoritmalarından faydalanarak model geliştirmeyi, incelemeyi ve uygulamayı amaçlamaktadır. Bu amaçla, üretim verileri, tarih-saat kayıtları, geçmiş güç üretimi ve sıcaklık ölçümleri kullanılmıştır. Tez kapsamında, EÜAŞ Aslantaş HES seçilmiştir. Güç üretim verileri, EXIST Şeffaflık Platformu'ndan alınmıştır. Veriler, Python kullanılarak işlenmiş, kodlanmış ve tahmin edilmiştir. Rastgele orman ve gradyan artırımlı karar ağaçları algoritmalarının uyumluluğu ve etkinliği incelenmiştir. En uygun algoritma, kapsamlı testlerle performansa bağlı olarak belirlenmiştir. Bu sistematik yaklaşım, hidroelektrik güç üretim tahmini için en uygun algoritmayı belirleyecektir. Analizler, rastgele orman ve gradyan artırımlı karar ağaçları modellerinin hidroelektrik üretimi tahmin etmede etkili olduğunu göstermiştir. Ancak, rastgele orman modeli daha üstün performans sergilemiştir.

Anahtar Kelimeler: Hidroelektrik, güç, tahmin, rastgele orman, gradyan artırımlı karar ağaçları.

ABSTRACT

**HYDROELECTRIC POWER FORECASTING
VIA TREE-BASED MACHINE LEARNING ALGORITHMS**

Bektaş Aykut ATALAY
M.Sc., Department of Electrical and Electronic Engineering

Supervisor: Asst. Prof. Dr. Kasım ZOR
June 2024, 48 pages

Hydroelectric power is the oldest and essential renewable energy source. Because energy demand constantly increases due to rapid economic and population growth in Türkiye, hydroelectric energy is crucial to Türkiye's energy resources. Hydroelectric plant planning and execution are pivotal for the state and energy enterprises. Hydroelectric power is proper for forecasting algorithms because of its seasonal dependency. Accurate forecasts of hydroelectric power provide carbon emission reduction, generation yields, and environmental sustainability. This thesis aims to develop, examine, and implement tree-based machine learning algorithms for forecasting power generation in an operational hydroelectric power plant with a capacity over 100 MW using generation data, date-time records, historical power generation, and temperature measurements. EÜAŞ Aslantaş HPP was selected to apply this thesis. The power generation data was fetched from the EXIST Transparency Platform. The data was processed, coded, and predicted using Python. Random forest (RF) and gradient boosted decision trees (GBDT) were examined for their compatibility and effectiveness. The most suitable algorithm was identified based on its performance through rigorous testing; this systematic approach will determine the optimal algorithm for hydroelectric power generation prediction. The analysis demonstrated that RF and GBDT models effectively forecasted hydroelectric power production. However, the Random Forest model showed slightly superior performance.

Keywords: Hydroelectric, power, forecasting, random forest, gradient boosted decision trees.

Dedication

To my life: Sultan, Zehra and Zeynep.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Asst. Prof. Dr. Kasım ZOR, whose unwavering support and encouragement have been invaluable throughout this journey. His insightful guidance, constructive feedback, and constant motivation have contributed significantly to my progress and success. I deeply appreciate his expertise and the time he dedicated to my work. In addition, I would like to show my gratitude to the jury members, namely Assoc. Prof. Dr. Necdet Sinan ÖZBEK and Asst. Prof. Dr. Oğuzhan TİMUR for their time, comments, and suggestions.

I would also like to extend my heartfelt thanks to my family, whose love and support have been my anchor. To my love, who has been a pillar of strength and patience, and to my children, who have been a source of joy and inspiration, thank you for standing by me and believing in me. This achievement is as much yours as it is mine.

Furthermore, I thank my colleagues and friends for their encouragement and understanding. Their support and fellowship have made this journey more enjoyable and fulfilling.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABC	:	Artificial Bee Colony
ANFIS	:	Adaptive Neuro-Fuzzy Inference System
ANN	:	Artificial Neural Network
ARIMA	:	Autoregressive Integrated Moving Average
BiLSTM	:	Bidirectional Long Short-Term Memory
CatBoost	:	Categorical Boosting
CNN	:	Convolutional Neural Network
DSİ	:	Devlet Su İşleri (The State Hydraulic Works)
EEMD	:	Ensemble Empirical Mode Decomposition
EMD	:	Empirical Mode Decomposition
EÜAŞ	:	Elektrik Üretim Anonim Şirketi (Electricity Generation Corporation)
EXIST	:	Energy Exchange Istanbul
GA	:	Genetic Algorithm
GBDT	:	Gradient Boosted Decision Trees
GES DISC	:	Goddard Earth Science Data Information and Services Center
GMAO	:	Global Modeling and Assimilation Office
HPP	:	Hydroelectric Power Plant
LightGBM	:	Light Gradient-Boosting Machine
LSSVM	:	Least Squares Support Vector Machine
LSTM	:	Long Short-Term Memory Network
MAE	:	Mean Absolute Error
MAM	:	Marmara Araştırma Merkezi (Marmara Research Center)
MAPE	:	Mean Absolute Percentage Error
MERRA	:	Modern-Era Retrospective Analysis for Research and Applications
NSE	:	Nash-Sutcliffe Efficiency
PCA	:	Principal Component Analysis
PSO	:	Particle Swarm Optimization
RCP	:	Representative Concentration Pathways
RF	:	Random Forest

RMSE	:	Root Mean Squared Error
RNN	:	Recurrent Neural Network
SARIMA	:	Seasonal Autoregressive Integrated Moving Average
SSA	:	Sparrow Search Algorithm
SVM	:	Support Vector Machine
TÜBİTAK	:	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (The Scientific and Technological Research Council of Türkiye)
XGBoost	:	Extreme Gradient Boosting

LIST OF SYMBOLS

hm^3	: Hecta cubic meter
kV	: Kilovolt
m	: Meter
m^3	: Cubic meter
MW	: Megawatt
MWh	: Megawatt hour
V	: Volt

1. INTRODUCTION

In this chapter, an introduction to the thesis with its primary goals, background and motivation, objectives, methodology are presented in order.

1.1. Background and Motivation

Hydroelectric power is an important renewable energy source, and it plays a significant role in global energy production. Together with power generation, flood control, irrigation, and water supply are its main benefits. It means agriculture, economy, protection of people and land against natural disasters, and clean water are linked to hydroelectric power, and vice versa as well.

Energy Info Bank of Türkiye shows that 36.3% of Türkiye's electricity is generated from coal, 21.4% from natural gas, 19.6% from hydraulic energy, 10.4% from wind, 5.7% from solar, 3.4% from geothermal and 3.2% from other sources. As seen in Figure 1.1, hydroelectric power shares 1 out of 5 in 2023 electricity power generation of Türkiye.

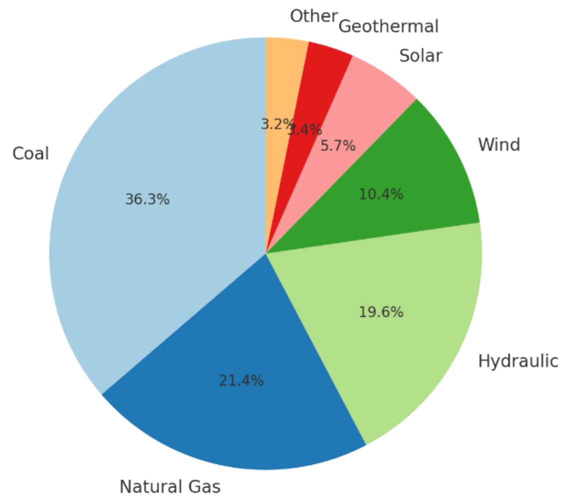

Figure 1.1. Türkiye's Electricity Generation Sources in 2023

In addition to its share in electricity generation, the cost of electricity generation and renewability are critical factors for Türkiye. As it uses the potential and kinetic energy in the water, only water and water control are required for hydroelectric energy production. The environment is not polluted as no other fuel is also needed.

Hydroelectric power generation is linked to many factors. Its generation is essential and priceless. Hydroelectric power generation should be maximised and maintained to get maximum benefits. For this purpose, hydroelectric power generation needs to be planned. This planning is done according to the supply and demand balance of the electricity market and seasonal conditions. For these reasons, the primary motivation of this thesis is forecasting hydroelectric power generation.

As the thesis aims to forecast hydroelectric power generation with tree-based machine learning (ML) algorithms, one hydroelectric power plant (HPP) is chosen. Aslantaş HPP is settled on the Ceyhan River. The Ceyhan River is one of the main waterways in the Eastern Mediterranean Region. Aslantaş HPP harnesses the Ceyhan River's energy to produce electricity. With a 138 MW installed capacity, Aslantaş HPP generates approximately 569 GWh of electricity annually with an average capacity factor of 47%. The vast reservoir created by the dam, with a capacity of 1,150 million m^3 at average water level, plays a crucial role in irrigation. It irrigates a substantial land area, estimated to be around 149,849 hectares. Aslantaş HPP is vital for managing floods along the Ceyhan River. The dam helps regulate water flow and mitigate potential flooding events. The construction of the Aslantaş Dam and HPP in the Ceyhan Basin began in 1976 and was completed in 1984. The dam, constructed with an earth-fill type structure, has a body volume of 8,493,000 m^3, a crest length of 585 m, and a height of 95 m. The dam's maximum water level is 155 m, storage capacity is 1,799 hm^3, and lake area is 63 km^2. The electrical energy generated at the power plant is levelled up to 154 kV and connected to the national electricity grid. (EÜAŞ Aslantaş HPP)

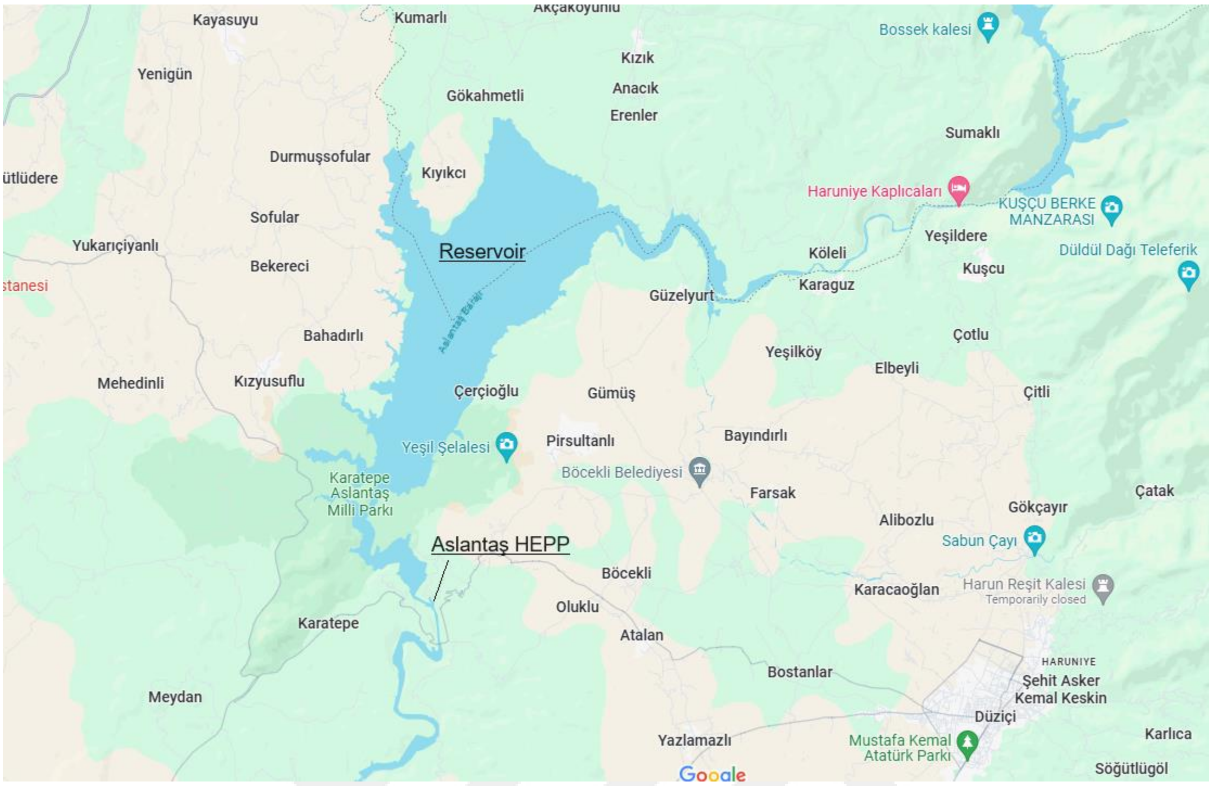

Figure 1.2. Aslantaş HPP Location

The motivation of the thesis is expressed as follows:

- Forecasting of hydroelectric power production is a crucial factor for optimising grid operations, ensuring a stable energy supply, and reducing operational expenses (Cassagnole et al., 2021).
- This thesis seeks to improve the precision of these forecasts to support more efficient electrical energy management.
- The variability of hydroelectric power generation and weather conditions underscores the importance of forecasting in Türkiye. While hydropower is the country's most widely used renewable energy source, its availability depends on rainfall, snowfall, and droughts. So, precise forecasting of hydroelectric power generation enables energy planners and operators to anticipate and mitigate the impacts of such variations, ensuring a reliable and stable electricity supply for the nation (Acaroğlu et al., 2023).

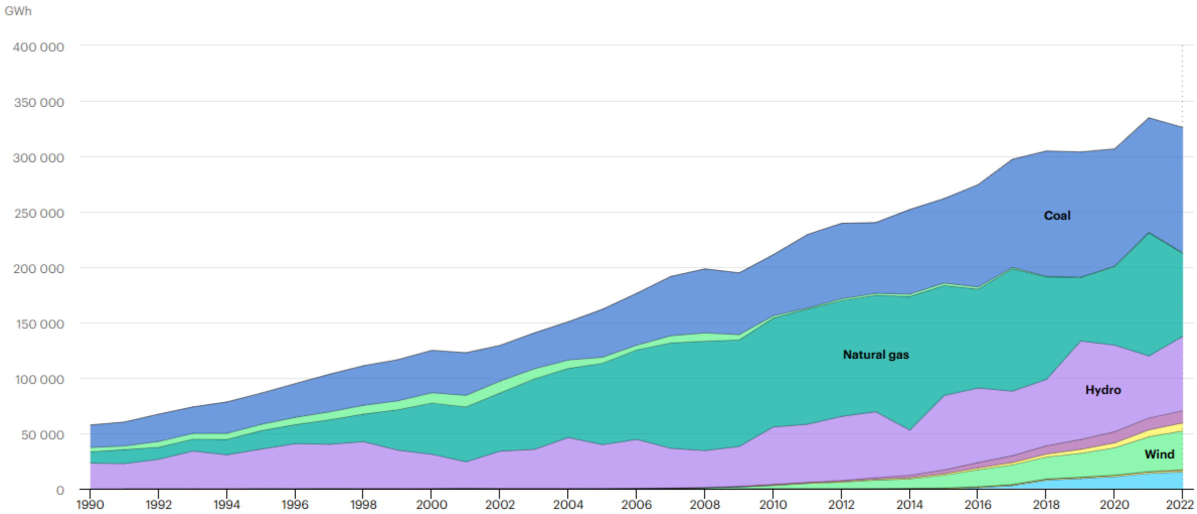

Figure 1.3. Electricity generation by source, Republic of Türkiye, 1990-2022. Source: IEA Energy Statistics Data Browser

- Since climate change and environmental sustainability are ongoing problems, better and more hydroelectric power production can reduce reliance on fossil fuels. This helps lower gas emissions and provides more sustainable energy sources.
- Better forecasting models can improve the financial planning of electricity and investment strategies for HPP projects. As forecasting generation accurately, maintenance schedules of plants and irrigation can be optimised, and water resources can be used more economically and effectively.
- HPPs are critical for controlling floods and managing water resources. Proper water flow and energy production forecasting help manage water management, reduce flood risks, and provide sufficient water for irrigation and tap water supply.
- Hydroelectric power is strategically essential in Türkiye, which has abundant water resources. The project "Flow Forecasting and Basin Optimization System for Dams and HPPs (ATHOM)" undertaken by DSİ and TÜBİTAK MAM, highlights this significance. This project aims to refine flow forecasts and basin management for hydroelectric dams and power plants. Proper flow forecasts and effective basin water management improve energy production efficiency and ensure the sustainable management of water resources.

- The thesis is looking for a contribution to renewable energy and artificial intelligence knowledge, potentially leading to further innovations in these fields.

1.2. Objectives

The main objectives of the thesis are described as follows:

- The main objective is forecasting hydroelectric power generation using tree-based ML algorithms.
- To determine variables for forecasting and find data sources for these variables.
- To define the critical variables affecting hydroelectric power generation. This objective requires analysing various factors such as water inflow, reservoir levels, weather conditions, and historical power generation data to determine the most critical variables for the model.
- The development of a forecasting model is central to this thesis. The model will be designed to predict hydroelectric power generation with high accuracy using tree-based ML algorithms. The focus will be on selecting and optimising algorithms that can handle the complexities and dynamics of hydroelectric power generation.
- To validate the effectiveness of the tree-based ML algorithms, one HPP will be chosen as a case study. This plant will provide the data necessary for training and testing the forecasting models. The plant selection criteria will include data availability, operational complexity, and representativeness of typical HPPs.
- Gathering meaningful and high-quality data is crucial for forecasting accurate models. From the chosen HPP and relevant external databases, the data are acquired on various parameters such as water inflow rates, reservoir levels, weather conditions, and historical power generation data. Cleaning and preprocessing the data to ensure it is suitable for analysis. This includes handling missing values, normalising data, and splitting the dataset into training and testing sets. Validating the quality and relevance of the dataset to ensure it is appropriate for training the forecasting models.
- To forecast electrical energy generation of an HPP by tree-based ML techniques containing gradient boosted decision trees (GBDT) and random forest (RF).

- To benchmark and understand the results using several key metrics: coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE).

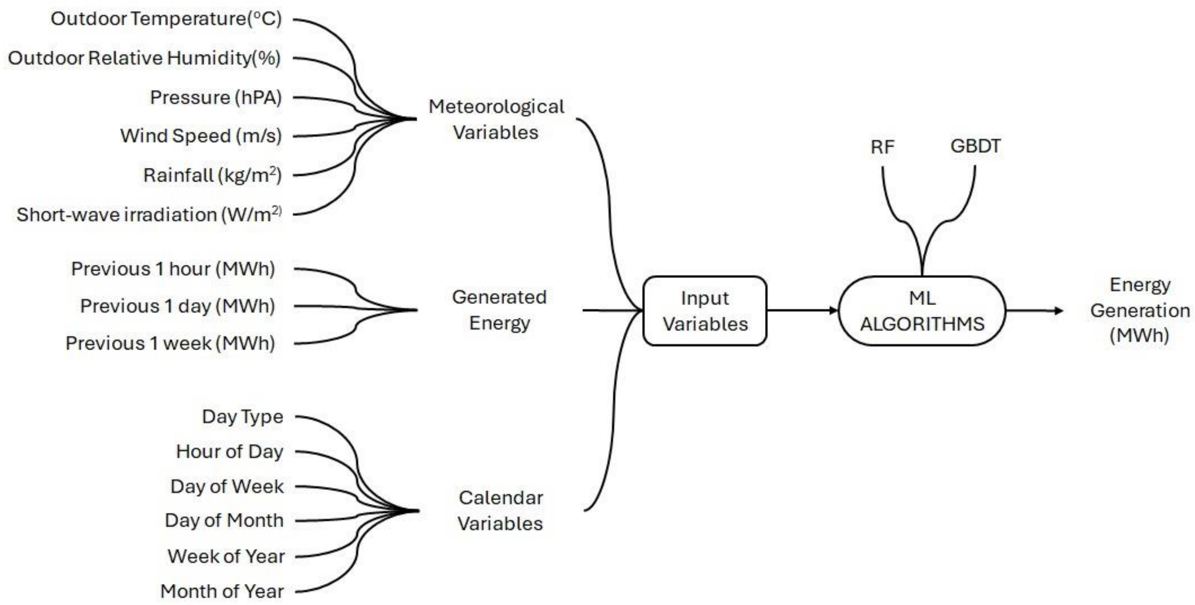

Figure 1.4. Steps of the Applied Methodology

1.3. Methodology

The steps of the applied methodology for the thesis are summarised in the previous figure given as follows:

- The input variable categories are energy generation, meteorological, and calendar variables.
- The target variable is real generated energy.
- Power generation data is obtained from the EPIAS Transparency Platform.
- 1-hour, 1-day, and 1-week lagged generation data is prepared in Python.
- Meteorological variables are taken from the MERRA-2 datasets (Global Modeling and Assimilation Office (GMAO) (2015)). It includes temperature, radiance, humidity, wind, rainfall, and short-wave irradiation. It was collected with a Python

script (Aiken, 2018). The GES DISC portal was used to fetch weather data for 2020, 2021, 2022, and 2023 respectively.

- Calendar variables include the month of the year, week of the year, day of month, hour of day, sample number of hours, and day type.
- After all energy generation, meteorological, and calendar variables are prepared, time stamps are equalled. As EPIAS Transparency Platform data's time zone is +3, MERRA-2 data's time zone is 0. In Python, MERRA-2 data's time column is shifted to 3 periods so that time zones are the same.
- Next, tree-based ML techniques, including GBDT and RF, were employed.
- Consequently, a benchmark analysis of the results achieved is presented meticulously by R^2, MAE, and RMSE

2. LITERATURE REVIEW

In this chapter, a summary of related literature, including master's and doctoral theses, journal and conference publications, which are contained by ProQuest Dissertations and Theses Database, Institute of Electrical and Electronics Engineers (IEEE) Xplore and ScienceDirect Digital Libraries, and other databases and libraries, are presented. Related journal and conference publications in the literature are mentioned. It should be noted that the literature is narrowed according to the following criteria:

- As the subject of the thesis, "hydroelectric power forecasting", "hydroelectricity power forecasting", "hydraulic power forecasting", and "hydropower forecasting" are selected for search.
- The search results are narrowed to "tree-based" and "machine learning."
- Next, the search was run using combinations of the terms respectively "hydroelectricity", "hydropower", "HPP"; "energy"; "power", "use", and "generation", "production"; "forecast", "forecasting", "prediction", "predicting", "estimation", and "estimating".

2.1. Related Theses

Demir (2015) proposed a method employing artificial neural networks (ANN) to estimate the energy production of Dim Dam in Türkiye. He developed a model using various ANN architectures to forecast the energy output based on input variables such as water level, volume, and discharge rates. He utilised historical data to train the networks and employed a backpropagation learning algorithm to adjust the weights and neuron connections according to changing conditions. He offered a solution for different operational scenarios of the dam, including varying water inflows and seasonal changes. Consequently, he compared his ANN-based model with a traditional energy production formula and multiple regression analysis for the same data. The proposed method provided lower errors, demonstrating its effectiveness in accurately predicting energy production. Demir's work highlights the potential of ANN in optimising hydropower generation by providing reliable forecasts and addressing the complexities of water resource management.

Dmitrieva (2015) explored several ML and statistical methods to forecast hourly energy production for two small hydropower plants in Hyen, Western Norway. Her study aimed to determine the best one-step-ahead prediction model for these plants, which have different hydrological characteristics due to their distinct locations. The first plant, located in the mountains, experiences immediate and short effects from rain, while the second plant, situated in a valley with a nearby pond, benefits from a more consistent water supply year-round. Dmitrieva compared methods, including ARIMA, regression analysis, neural networks, and support vector machines (SVMs). She found that ML models, particularly neural networks and SVMs, outperformed traditional statistical models in capturing the complex, nonlinear relationships inherent in the data. Her results indicated that these advanced models provided more accurate and reliable forecasts for the energy production of the small hydropower plants, demonstrating significant improvements in prediction accuracy and operational efficiency. The study highlighted the importance of selecting appropriate forecasting models tailored to each hydropower plant's specific characteristics and data patterns.

Afonso De Sousa (2019) proposed a method using various statistical models to forecast electricity generation for small hydropower plants (PCHs) in Brazil. The study, conducted as a master's thesis at PUC-Rio under the supervision of Reinaldo Castro Souza, focused on using causal models like the Box and Jenkins transfer function and dynamic regression and univariate models like Holt-Winters exponential smoothing and ARIMA. The primary aim was to improve the accuracy of generation forecasts, particularly in the absence of regulation reservoirs, by incorporating inflow data from neighbouring hydro plants as an exogenous variable. The study analysed historical data from several PCHs, applying spatial autocorrelation to identify the best-neighbouring inflow series for improving forecast accuracy. The results demonstrated that causal models incorporating neighbouring inflow data outperformed univariate models regarding forecast accuracy. Specifically, the Box and Jenkins transfer function and dynamic regression models provided better predictions for the small hydropower plants by effectively capturing the influence of nearby hydrological conditions. The improved forecasts, with reduced mean absolute percentage error (MAPE) and root mean square error (RMSE), suggest practical applications for managing PCH operations and planning in the Brazilian energy sector. The study highlights the importance of utilising spatial and hydrological data to enhance the reliability of energy production forecasts, ensuring more stable and efficient power generation from renewable sources.

Mnguu (2020) proposed a method which employs the ARIMA model to forecast hydro-power generation at Mtera Dam in Tanzania. He compared several time series models in his work, including SARIMA, SVM, and ANN. The study identified ARIMA as the most adequate model for the data, achieving a high prediction precision. The nonlinear relation of electric net power generation was explored using historical monthly recorded data. Experimental results demonstrated that the ARIMA model provided suitable predictions, and TANESCO was recommended to use the model for operational and planning activities. The proposed method provided accurate forecasts with a MAPE of 3.12%, ensuring the stability and efficiency of the power grid. The research utilised historical monthly recorded data to explore the nonlinear relationship between various factors affecting electric net power generation. Through extensive data analysis, Mnguu found that the ARIMA model outperformed the other models regarding prediction accuracy. The ARIMA model was particularly effective in capturing the seasonal patterns and trends in the data, making it the most adequate model for the dataset. Experimental results demonstrated that the ARIMA model provided accurate forecasts, with a Mean Absolute Percentage Error (MAPE) of 3.12%. This level of precision in prediction ensured the stability and efficiency of the power grid.

Wergeland (2023) proposed an ML-based method to forecast hourly hydropower production for the Bjørgum and Furegardane run-of-river power plants in Norway. This study aimed to enhance the accuracy of next-day production forecasts to ensure better grid stability and minimise imbalance fees. Wergeland compared the performance of three ML models: a random forest regressor, a multilayer perceptron neural network (MLP), and a long short-term memory (LSTM) neural network. The methodology involved extensive data preprocessing and feature engineering, using historical production data from 2017 to 2022, weather forecasts and observational data. Features included accumulated rainfall and snowmelt, which are critical for capturing the hydrological inputs affecting power generation. The study found that the LSTM neural network outperformed the random forest regressor and the MLP neural network in predicting hourly production for both power plants. The LSTM model's ability to capture long-term dependencies in time series data resulted in more accurate forecasts, significantly improving Småkraft's. For Bjørgum, the LSTM model achieved lower mean absolute error (MAE) and mean squared error (MSE) compared to the other models and the existing forecasting approach. Similarly, for Furegardane, the LSTM model provided the best predictions, demonstrating its robustness across different hydrological conditions. Wergeland also identified the most critical weather variables for making accurate predictions, highlighting the practical application of these models for enhancing the stability and efficiency of the power grid.

2.2. Related Journal and Conference Publications

Joaquim et al. (2005) used ANN for temporal processing in the context of electric power generation. They utilised historical time-series data to train their models, focusing on capturing seasonal and temporal patterns. The results demonstrated that ANNs could effectively model complex temporal relationships in power generation data, providing accurate short-term forecasts. (Joaquim et al., 2005).

S. Wang et al. (2011) Propose a seasonal decomposition-based least squares support vector machine (LS-SVM) model using historical power demand data to decompose the time series into seasonal components before applying the LS-SVM for forecasting. The results showed that this method improved prediction accuracy by effectively capturing seasonal variations, making it a valuable tool for power system operators.

Monteiro et al. (2013) presented an original short-term forecasting model for electric power systems based on adaptive neuro-fuzzy inference systems (ANFIS) using a dataset of historical power system variables, including load, temperature, and humidity. The results demonstrated that the ANFIS model could adapt to changing conditions and provide accurate short-term forecasts, highlighting its potential for improving power system reliability and efficiency.

Aleksandrovskii and Borshch (2013) used a multiple linear correlation with stepwise regression analysis to predict electric power generation at Rybinsk HPP and Nizhegorodskaya HPP using historical generation data, hydrological variables, and weather. The findings indicated that the selected statistical methods could provide reliable predictions of hydropower generation, assisting in better operational planning and resources.

Uzlu et al. (2014) proposed a method including the ANN (artificial neural network) model with the ABC (artificial bee colony) algorithm to estimate annual hydraulic energy production of Türkiye using energy demand, energy consumption, temperature, and population to estimate hydroelectric generation in Türkiye using neural networks. The results showed that neural networks could effectively predict hydroelectric generation, with a high correlation between predicted and actual values.

G. Li et al. (2014) developed a short-term power generation energy forecasting model for small hydropower stations using a Genetic Algorithm - Support Vector Machine (GA-SVM) approach. The model's performance was benchmarked against the ARMA model, showing superior performance and effectiveness in forecasting short-term power generation energy for small hydropower stations.

Monteiro et al. (2014) proposed a short-term forecasting model for aggregated regional hydropower generation. The H4C2 model utilised past recorded hourly electric power values, hourly water precipitation forecasts, load demand, and wind generation. The model achieved satisfactory next-day forecasts for real-life scenarios in Spain and Portugal. (Monteiro et al., 2014)

G. Li et al. (2015) applied a correlation analysis method to long-term power production forecasting at small hydropower plants. The method achieved a prediction accuracy of 87.9% and an average relative error of 18.5%

G.-D. Li et al. (2016) enhanced the traditional grey model (GM(1, 1)) by integrating it with a regression model and applying the Markov chain to improve prediction accuracy. The model was validated with data from 2010 to 2015, showing its effectiveness in predicting hydroelectric power generation in Japan, although the growth rate remained relatively stable.

M. Li et al. (2016) addressed the unpredictability in small hydropower generation by combining wavelet decomposition with a backpropagation (BP) neural network optimised by Particle Swarm Optimization (PSO). This method decomposed historical load data to enhance prediction accuracy. Applied to a small hydropower-rich region, the approach achieved a prediction precision of 93.7%, significantly improving over traditional high-voltage system criteria.

Hussin et al. (2016) proposed a hybrid metaheuristic approach combining ANN with the Bat Algorithm (BA) to forecast electricity production and water consumption at Sultan Azlan Shah Hydropower Plant. The hybrid ANN-BA model demonstrated superior performance over conventional ANN models, effectively handling time series input data and assumptions specific to the study.

Fan Qiang et al. (2016) developed a short-term power forecast system for run-of-river small hydropower stations (RSHSG) in Guizhou, China. The system, tested with April and May 2016 data, showed significant improvements in grid dispatch management and operational safety. The system's application in the Guizhou Power Grid underscored its practical utility and effectiveness.

Konica and Staka (2017) applied fuzzy time series techniques to forecast energy production in small hydropower plants in Albania. This method effectively handled vague and incomplete data, providing better accuracy and reduced computational overhead compared to other models.

Razi et al. (2017) focused on predicting power generation in small hydropower systems at Sungai Perting Bentong, Malaysia, using ultrasonic level sensors to measure water flow rates. Their study demonstrated that the power output exceeds 1 MW and is highly dependent on water flow fluctuations. This model is helpful for daily and monthly monitoring of system efficiency.

Z.-X. Wang et al. (2017) proposed a data grouping approach based on the grey model GM(1,1) to forecast quarterly hydropower production in China. By dividing the time series into seasonal groups, their method improved prediction accuracy, with a mean absolute per cent error (MAPE) of 16.2%, outperforming traditional GM(1,1) and SARIMA models.

Kiruthiga and Amudha (2018) saw advancements in hydropower generation optimisation and forecasting using Particle Swarm Optimization (PSO). This approach aimed to optimise the generation process and improve forecasting accuracy, though the available literature did not detail specific metrics and results.

Hammid et al. (2018) utilized a feed-forward back-propagation ANN to predict power production at the Himreen Lake Dam in Iraq. Their model, trained on 3570 data points over ten years, achieved a correlation coefficient (R) above 0.96 between predicted and observed outputs, demonstrating high accuracy in performance predictions.

Dehghani et al. (2019) presented the prediction of hydropower generation using a Grey Wolf Optimization Adaptive Neuro-Fuzzy Inference System (GWO-ANFIS). This method combines Grey Wolf Optimization with an Adaptive Neuro-Fuzzy Inference System to enhance forecasting accuracy.

Lopes et al. (2019) compared polynomial neural networks and conventional ANNs to predict hydropower generation potential in the Amazon region. This case study demonstrated the applicability of ANN in diverse geographical contexts, though specific results and error metrics were not mentioned. Their study found that the group method of data handling (GMDH) slightly outperformed conventional ANN models, indicating its usefulness for energy planning.

L. Li et al. (2019) proposed a deep neural network approach combining residual and recurrent neural networks for hydroelectric power generation prediction. Their model, which integrated multiple data granularity levels, significantly improved prediction performance, showcasing the potential of deep learning in industrial applications. The model incorporated multi-grain hydropower generation data divided into recent, daily, weekly, and time-series sequences and used a multi-information fusion method to integrate these predictions.

Barzola-Monteses et al. (2019) developed an ARIMAX model incorporating past generated energy and precipitation data to predict hydroelectric power production in Ecuador. This study utilised Box-Jenkins and Box-Tiao time series analysis techniques, specifically the ARIMA and ARIMAX models, to predict hydroelectric power production in Ecuador.

de Sousa et al. (2020) used time series models correlating future inflow and past generation data to forecast electricity generation for Brazil's small hydropower plants (SHPs). The methodology showed reduced forecasting errors, providing a reliable tool for electricity distribution companies to estimate future electricity generation accurately.

Zhou et al. (2020) introduced DeepHydro, a novel stochastic method for modelling multivariate time series and forecasting power generation in hydropower stations. It uses conditioned latent recurrent neural networks to capture temporal dependencies and uncertainties in co-evolving time series data, such as temperature, outflow, and inflow. The model employs continuous normalising flow and neural ordinary differential equations to handle discrete observations within continuous dynamic systems.

Lian et al. (2020) proposed a trend-guided extreme learning machine (TG-ELM) model to improve the accuracy of hydropower generation forecasts for small hydropower groups in Guangxi. The model enhances traditional extreme learning machines by incorporating trend features extracted from smoothed and repaired historical data. Based on August 2019 data, simulation analysis showed that TG-ELM achieved a root mean square error of 4.16 MW and an average absolute error of 6.11%, outperforming backpropagation neural networks and support vector machines in accuracy and computation time.

Javed et al. (2020) developed a forecasting model for electricity generation at Tarbela Dam in Pakistan, incorporating temperature and rainfall data from the dam's catchment area. The model applied various supervised ML and time-series methods to forecast energy production, achieving a mean absolute error of 2.47. This approach helps policymakers manage the gap between energy demand and production and detect anomalies in electricity generation.

Ahmad and Hossain (2020) explored the optimisation of reservoir operations based on short-term inflow forecasts derived from numerical weather prediction (NWP) models. The study used forecast fields from the Global Forecast System (GFS) and the Variable Infiltration Capacity (VIC) hydrologic model to predict reservoir inflows for 1–16 days. The optimisation approach significantly improved hydropower generation at two US dams without compromising flood control and dam safety.

Zhong et al. (2020) combined a macro-scale distributed hydrological model with an optimal operation model to predict future hydropower generation in the upper Yangtze River basin under climate change scenarios. The study covered 62 reservoirs and used the Statistical Downscaling Model (SDSM) to predict future rainfall and temperature trends. Results indicated no significant change in hydropower generation under the RCP2.6 scenario, while the RCP4.5 and RCP8.5 scenarios showed increasing trends, particularly under RCP8.5, highlighting the sensitivity of hydropower generation to climate change.

Rahman et al. (2021) reviewed ML-based approaches, particularly ANN models, for predicting renewable energy generation, including hydropower. The study emphasised the use of various ANN architectures such as multi-layer perceptron (MLP), recurrent neural network (RNN), convolutional neural network (CNN), and long short-term memory (LSTM) for short-term time-series predictions. These models are particularly suitable for processing complex time series inputs in hybrid renewable energy systems, offering accurate and reliable predictions.

S. Yang et al. (2021) developed a multimodal deep learning method combining CNN and MLP to forecast daily power generation for small hydropower stations in South China. By considering the temporal and spatial distribution of precipitation, the model achieved a prediction accuracy of 93%, outperforming support vector regression (SVR) and LSTM methods. This approach improved accuracy by approximately 5.8% compared to models that did not account for spatial precipitation distribution.

Sessa et al. (2021) applied Random Forest models to forecast run-of-river hydropower generation based on climate variables across Europe. The study demonstrated that finer spatial resolution of climate data significantly improved model accuracy, highlighting the importance of precise climate data in hydropower forecasting.

Al Rayess and Ülke Keskin (2021) utilised various ML techniques, including Decision Tree, Deep Learning, Generalized Linear, Gradient Boosted Trees, and Random Forest models, to forecast hydropower generation at Almus Dam in Türkiye. The Gradient Boosted Trees model achieved the highest correlation value, indicating its effectiveness for hydropower prediction.

Ekanayake et al. (2021) developed regression-based models using rainfall, temperature, and evaporation data to predict power generation at the Samanalawewa plant in Sri Lanka. Gaussian process regression (GPR) and support vector regression (SVR) provided the most accurate predictions, with GPR producing the best correlation coefficient closer to 1 with minimal error (Ekanayake et al., 2021).

Polprasert et al. (2021) proposed a long-term forecasting model using the autoregressive integrated moving average (ARIMA) method for the Son La hydropower plant in Vietnam. The model demonstrated an upward trend in future electricity generation, providing a useful tool for strategic energy planning and addressing environmental variability.

Yildiz and Açikgöz (2021) proposed a hybrid method combining extreme learning machine (ELM) with the artificial bee colony (ABC) algorithm to forecast small hydropower plant generation. This approach improved prediction accuracy by optimising the input weights of the ELM, showing significant advantages over traditional methods.

Zolfaghari and Golabi (2021) combined adaptive wavelet transform (AWT), LSTM, and Random Forest (RF) algorithms to predict electricity production in hydropower plants. The hybrid model outperformed benchmark models such as ARIMA-GARCH and wavelet transform-feed forward neural networks, demonstrating superior predictive power for non-stationary and nonlinear EP series.

Z. Yang et al. (2021) developed a framework for short-term hydropower generation scheduling that accounts for uncertainties in load demand forecasts. The framework utilized a martingale model of forecast evolution (MMFE) and an enhanced shuffled frog leaping algorithm (SFLA) to solve scheduling problems and quantify risks efficiently.

Ren et al. (2021) proposed the DWHO algorithm to optimize water resource management and predict hydropower generation. The algorithm demonstrated superior convergence speed and accuracy, producing about 17% more electricity than others, with high reliability and resilience indices.

Huangpeng et al. (2021) the DCSA algorithm combined with an ANN model was used to predict future hydropower generation under climate change scenarios. The model showed a decrease in annual power generation under RCP2.6, RCP4.5, and RCP8.5 scenarios, with the highest accuracy among compared algorithms.

Barzola-Monteses et al. (2022) applied ANN models to predict short and medium-term hydroelectric production in Ecuador, considering historical production and precipitation data. The MLP univariate model outperformed other ANN architectures, offering a valuable tool for energy planning and decision-making (Barzola-Monteses et al., 2021).

Sarpong and Agyei (2022) applied an Autoregressive Integrated Moving Average (ARIMA) model to forecast the water level and power generation at Ghana's Akosombo dam. The study predicted a decrease in power generation from 398 MW/hr in December 2019 to approximately 374 MW/hr by December 2022. Similarly, the water level was forecasted to decline from 264.8 ft to approximately 255.19 ft by the same period. The study emphasises the need for proactive measures to expand renewable energy sources.

Xia et al. (2022) investigated forecasting models for small hydropower stations. Although the abstract did not provide specific model details, the study contributes to understanding and improving power generation forecasts for small-scale hydropower facilities.

Drakaki et al. (2022) developed uncertainty-aware forecasts for day-ahead energy production in small hydropower plants (SHPPs) without storage capacity. The study found that incorporating expert knowledge about hydrological regimes and SHPP technical characteristics within model training significantly improved forecasting accuracy.

J. Wang et al. (2022) proposed a fusion model combining Ensemble Empirical Mode Decomposition (EEMD), Adaptive Moment Estimation (ADAM), and Gated Recurrent Unit (GRU) neural networks. This model outperformed traditional forecasting methods, providing accurate forecasts of China's monthly hydropower generation and estimating economic benefits.

Rathnayake et al. (2022) utilised a Cascaded Adaptive Network-Based Fuzzy Inference System (ANFIS) to forecast hydropower generation at the Samanalawewa Reservoir, accounting for climate change impacts. The ANFIS algorithm performed superior to other state-of-the-art regression models, with minimal error rates.

Ji et al. (2022) developed a multi-model fusion approach combining GBDT, extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM) for short- and medium-term power demand forecasting. The XLG-LR fusion model outperformed neural network models such as GRU, LSTM, and TCN in accuracy and computational efficiency, making it a reliable tool for power enterprises and future power construction planning.

Corbari et al. (2022) presented a system that integrates numerical weather prediction models with snow melting and accumulation models to forecast hydropower production in mountainous areas. Despite challenges in providing reliable quantitative weather forecasts in complex alpine regions, this system proved effective in simulating lake inflow and energy production.

Bilgili et al. (2022) applied a Long Short-Term Memory (LSTM) neural network for one-day-ahead energy production forecasting from run-of-river hydroelectric plants in Türkiye. The LSTM model demonstrated higher accuracy than other data-driven methods like ANFIS with different clustering techniques, achieving a mean absolute percentage error (MAPE) of 5.98%.

Prakash et al. (2022) utilised statistical analysis and ML models to predict energy generation at the Tedzani hydropower plant in Malawi. This study demonstrated the potential of ML in optimising energy harnessing and consumption.

Velasquez and Flores (2022) proposed deep learning models for predictive maintenance at the Peña Blanca HPP. Their models, including a Deep Neural Network and an LSTM with an Autoencoder, effectively detected anomalies, enhancing maintenance efficiency and reliability.

Safaraliev et al. (2022) addressed medium-term forecasting (MTF) of power generation by hydropower plants in isolated power systems, incorporating atmospheric parameters such as air temperature, wind speed, and humidity. The study compared various ML models and found the Adaptive Boosting Linear Regression model to be the most straightforward and reliable for tasks with limited training samples (Safaraliev et al., 2022).

Kumar et al. (2022) proposed using ANN to predict energy generation targets for individual hydropower plants. The study achieved a correction coefficient higher than 0.99, utilising yearly energy generation data from multiple plants, demonstrating the reliability and predictability of hydropower as an energy source.

S. Zhang et al. (2022) introduced a hybrid model combining improved ensemble empirical mode decomposition (EEMD) and adaptive switching among artificial neural networks (ANNs) to predict generation at Pubugou Hydropower Station in Sichuan, China. This approach reduced prediction complexity and improved accuracy by switching between extreme learning machine (ELM), backpropagation neural network (BP), and general regression neural network (GRNN) based on performance. The model demonstrated vital forecasting accuracy and computational efficiency.

Zeng et al. (2023) developed a novel grey combination optimisation model based on the three-parameter discrete grey model TDGM (1,1) to predict China's hydropower generation. The model forecasts a 24.5% increase in hydropower generation by 2025 compared to 2020. The study highlights the optimisation algorithm and its implementation process, offering solutions to the low accuracy of traditional models in China's hydropower prediction.

İnal et al. (2023) developed a Linear Weighted Normalized Radial Basis Function (LWNRBF) neural network to estimate next-day energy generation at Darıca-2 Hydropower Station in Türkiye. Using reservoir inflow rates, current water levels, and weather forecasts, the model achieved average training and testing error percentages of 0.0012% and -0.0044%, respectively, indicating high accuracy.

Zou et al. (2023) proposed a prediction method combining SSA, PCA, EMD, and LTSM. This method, applied to a hydropower station in Sichuan, China, effectively reduced random fluctuation and improved prediction accuracy through optimised model parameters.

H. Zhang et al. (2023) developed an EEMD-LSTM model to forecast hydropower generation by considering the coupling effects of adjacent plants and multidimensional Meteorological factors. Applied to a basin hydropower plant, the model significantly improved prediction accuracy by integrating data from neighbouring.

Cheganova et al. (2023) used ARIMA models to forecast water inflow to the Sayano-Shushenskaya HPP reservoir. The k-nearest neighbour model, with a depth of retrospective data of 24 months, yielded the best results, achieving a determination coefficient of 0.749.

Sattar Hanoon et al. (2023) investigated various ML algorithms, including artificial neural networks (ANN), ARIMA, and support vector machines (SVM), to predict power production at the Three Gorges Dam. The models were effective across different prediction scenarios (daily, monthly, and seasonal), with high accuracy and reliable uncertainty measures.

Achig et al. (2023) utilised LSTM neural networks to predict the flow necessary to meet energy demands. This model achieved an absolute error of 5.12%, significantly contributing to Ecuador's National Development Plan for optimising water resource use.

Çakır (2023) proposed a hybrid methodology combining intuitionistic fuzzy time series and fuzzy c-means methods for renewable energy generation forecasting in Türkiye. The model showed high accuracy, making it applicable to various time series forecasting problems.

Z. Li et al. (2023) introduced a novel model using the Weighted Average Weakening Buffer Operator based on the Fractional order accumulation Seasonal Grouping Grey Model (WAWBO-FSGGM(1,1)). This model achieved the lowest MAPE values across multiple regions, indicating superior predictive performance.

Tebong et al. (2023) developed a two-level deep learning ensemble model for short-term hydropower forecasting, incorporating LSTM, Conv1D, and dense neural networks. This model outperformed individual deep learning models, demonstrating high predictive accuracy with an NSE of 0.702.

Ehtearm et al. (2023) proposed a hybrid model combining CNN-ANN with Gaussian Process Regression and the Salp algorithm for predicting hydropower production. This approach enhanced forecasting accuracy by integrating deep learning with regression techniques.

Abu et al. (2024) employed the Box-Jenkins model to forecast hydropower production at Tasik Kenyir, Terengganu. This traditional time series approach provided reliable predictions for hydropower generation.

Wei et al. (2024) introduced a transfer learning method for predicting power generation in data-scarce areas by integrating public knowledge from run-off small hydropower (RSHP) databases. The method involved a CNN-BiLSTM hybrid pre-trained network and hyper-parameter fine-tuning, significantly improving prediction accuracy with an average error reduction of 16.27% compared to traditional models.

3. MATERIALS AND METHODS

3.1. Data

- The input variable categories are energy generation, meteorological, and calendar variables.
- The target variable is real generated energy.
- The historical electrical energy generation variables are data from the previous 1 hour, the previous day (the same time in the previous day), and the previous week (the same time and day in the previous week).
- The period selected for input data collection is from 01.01.2020 to 01.01.2024.
- Power generation data was taken from the EPIAS Transparency Platform. The data covers 2020, 2021, 2022, and 2023, with hourly records for each year.
- Meteorological variables are taken from the MERRA-2 datasets (Global Modelling and Assimilation Office (GMAO) (2015)). It includes temperature, radiance, humidity, wind, rainfall, and short-wave irradiation. It was collected with a Python script (Aiken, 2018). The GES DISC portal was used to fetch weather data for 2020, 2021, 2022, and 2023.
- Month of the year, week of the year, day of month, hour of day, number of hours sampled, and type of day are calendar variables.

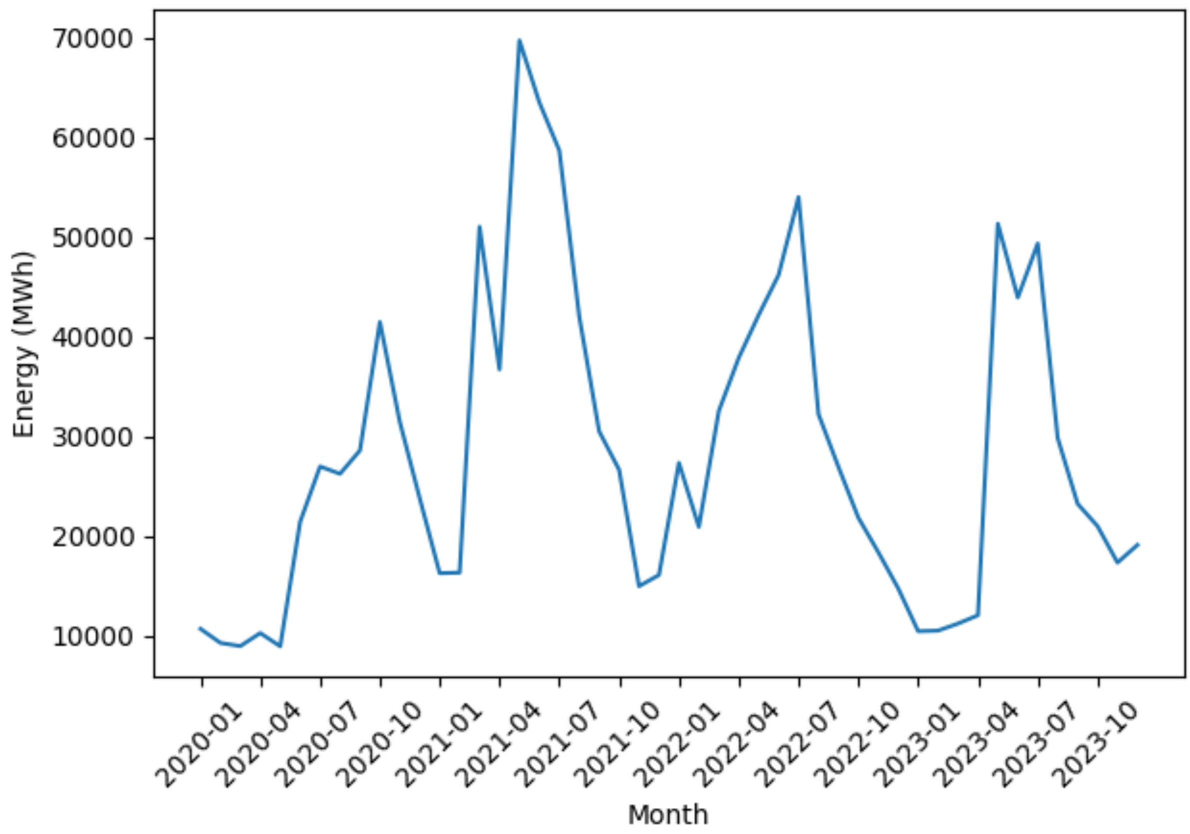

Figure 3.1. Monthly Generated Data

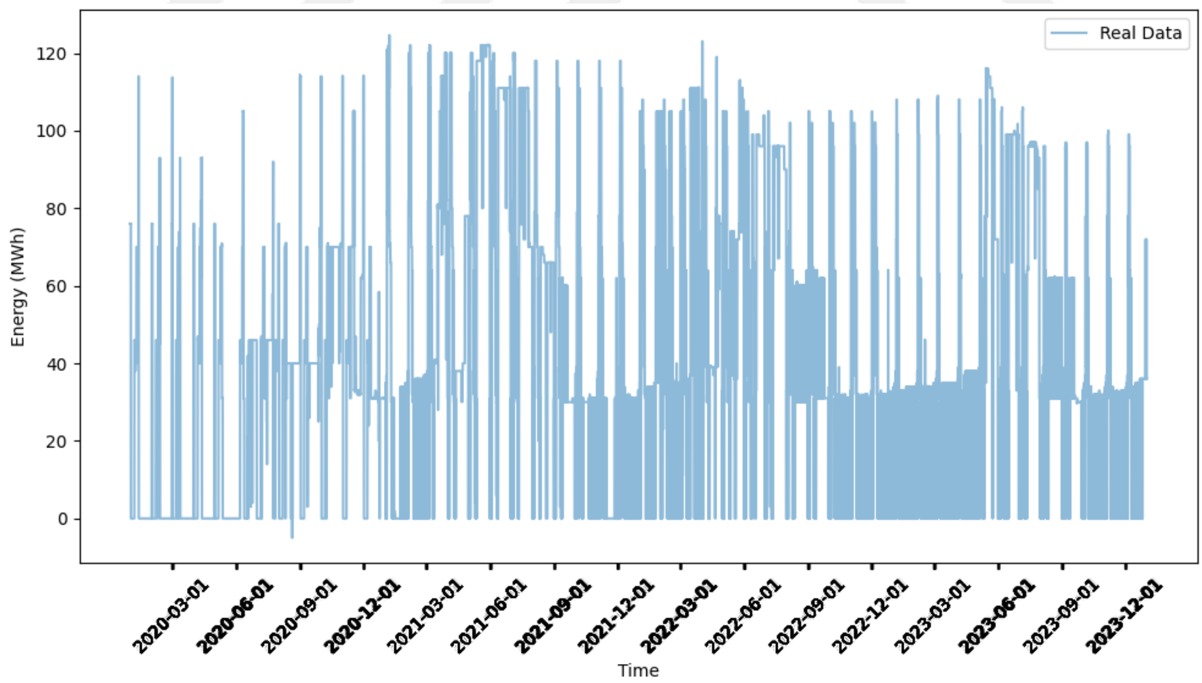

Figure 3.2. Hourly Data

The figures illustrate the trends and patterns in the generated energy data, visually confirming the quantitative evaluation. Figure 3.1 shows the overall generated energy, while Figure 3.2 details the monthly variations, highlighting seasonal dependencies and the impact of varying inflow conditions.

3.2. Tree-Based Machine Learning Algorithms

In this section, tree-based ML algorithms, including GBDT and RF, are expressed in detail.

3.2.1. Random Forests

RF construct multiple decision trees during training and combines their outputs to improve predictive performance and robustness. This method is known for its accuracy, ability to handle high-dimensional data, and minimal parameter tuning requirements. RF are ensembles of decision trees where each tree is built using a random subset of features and data samples. Predictions are made by averaging the outputs of individual trees (Breiman, 2001; Scornet et al., 2015).

Bagging involves generating multiple training datasets by random sampling with replacements from the original dataset. Each of these datasets is used to train a separate decision tree. The final prediction is made by aggregating the predictions of all the individual trees, usually through majority voting for classification or averaging for regression. At each split in the tree, a random subset of features is selected, and the best split is chosen only from these features. This introduces diversity among the trees and helps reduce the correlation between them (Biau and Scornet, 2016).

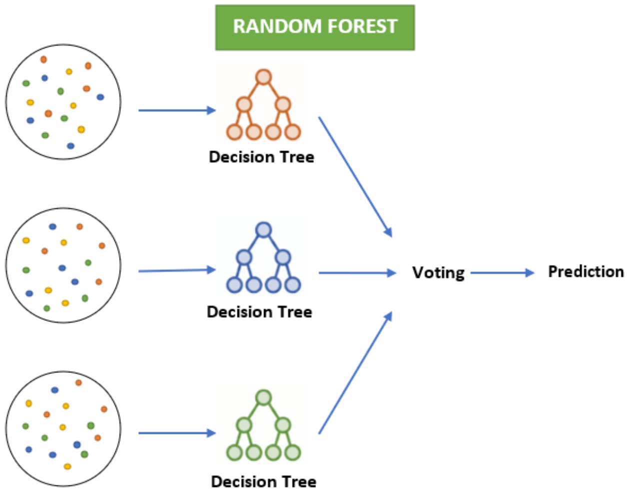

Figure 3.3. An illustration of RF (Luwe et al., 2023)

RF are composed of numerous decision trees that are trained on various subsets of the data. The prediction for a single tree in the forest is given by:

$$m_n(x; \Theta_j, D_n) = \sum_i \frac{\mathbb{1}_{X_i \in A_n(x;\Theta_j,D_n)} Y_i}{N_n(x; \Theta_j, D_n)} \tag{3.1}$$

where Θ_j represents the random parameters for the j-th tree, $D_n(\Theta_j)$ is the subset of data used for the j-th tree, $A_n(x; \Theta_j, D_n)$ is the cell containing x, and $N_n(x; \Theta_j, D_n)$ is the number of points in $A_n(x; \Theta_j, D_n)$, i *is* an element of set $D_n(\Theta_j)$.

Each tree predicts the average of the Y_i values in the leaf node containing x. The overall prediction of the RF is obtained by averaging the predictions of the individual trees:

$$m_{M,n}(x; \Theta_1, \ldots, \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^{M} m_n(x; \Theta_j, D_n) \tag{3.2}$$

where M represents the number of trees in the forest, $\Theta_1, \ldots, \Theta_M$ is the random variables governing the construction of the M trees and $m_n(x; \Theta_j, D_n)$ is the prediction of the j-th tree for input x.

As the number of trees M tends to infinity, the forest prediction converges to the expected prediction over the random parameters:

$$m_{\infty,\mathrm{n}}(x; D_n) = E_{\Theta}[m_n(x; \Theta, D_n)] \tag{3.3}$$

3.2.2. Gradient Boosted Decision Trees

GBDT is a robust algorithm that combines multiple decision trees to form a strong predictive model. The method involves building trees Sequentially, with each new tree attempting to correct errors made by the previous ones. This approach enhances the model's performance by focusing on the most difficult-to-predict cases. GBDT has become one of the most popular and practical algorithms for various ML tasks, including classification and regression.

Friedman developed the method of gradient boosting by applying the gradient descent technique to optimise loss functions, leading to the creation of GBDT (Friedman, 2001). Over the years, various improvements and extensions have been proposed, such as XGBoost, LightGBM, and CatBoost, each optimising different aspects like computational efficiency and handling of categorical features (Coadou, 2022).

The mathematical foundation of GBDT lies in gradient descent optimisation. The algorithm iteratively adds decision trees to minimise a specified loss function. Each tree is built to fit the residual errors (the differences between the observed and predicted values) of the combined ensemble of trees from previous iterations. The new trees are constructed to predict the negative gradient of the loss function concerning the current model's predictions. This process continues until the model reaches a predetermined number of trees or an acceptable level of performance. (Z. Zhang and Jung, 2019).

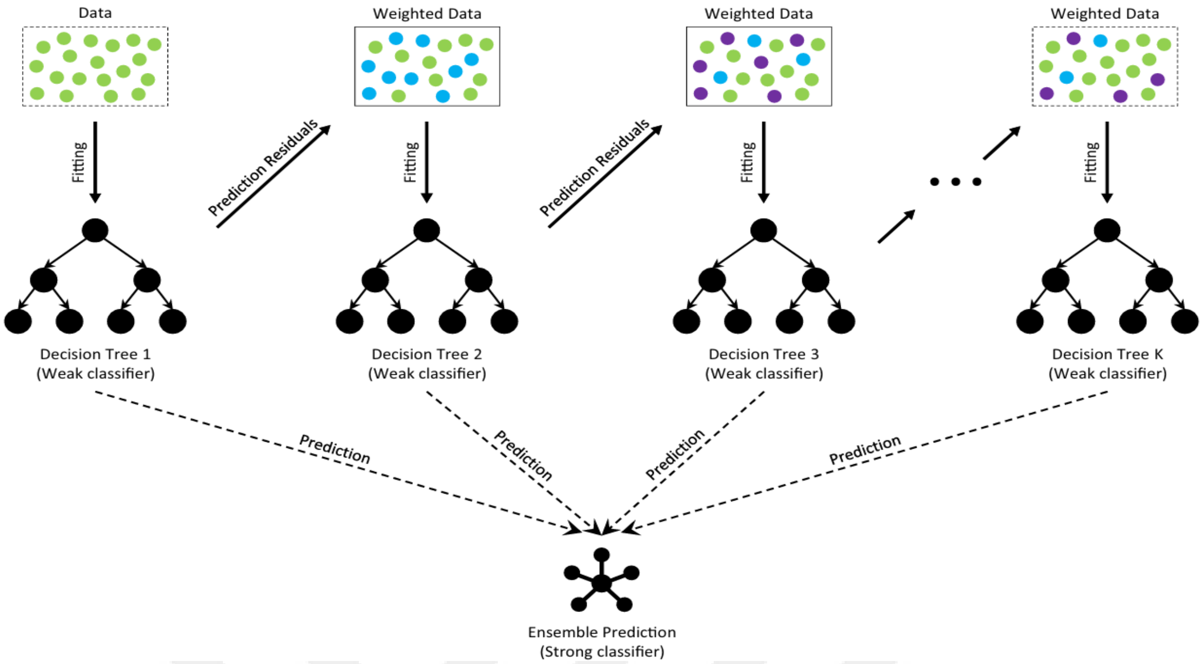

Figure 3.4. The Architecture of GBDT (Deng et al., 2021)

GBDT are an advanced form of boosting algorithms that combine the predictions of multiple decision trees to improve the overall predictive performance. Unlike AdaBoost, which adjusts the weights of misclassified instances, GBDT use gradient descent to minimise a loss function. GBDT build the model in a stage-wise fashion. At each stage, a new tree is added to the ensemble to correct errors made by the existing model (Coadou, 2022; Zor and Çelik, 2021).

GBDT are formulated as follows:

- The model F(x) is built iteratively by adding a new tree $h_k(x)$ to improve the prediction:

$$F_m(\mathrm{x}) = F_{m-1}(x) + h_m(x) \tag{3.4}$$

- The process begins with an initial model $F_0(x)$ which can be a constant value:

$$F_0(x) = argmin_c \sum_{i=1}^{n} L(y_i, c) \tag{3.5}$$

where $L(y_i, c)$ is the loss function, and y_i are the target values, $argmin_\gamma$ denotes finding the value of c that minimises the sum of the loss function overall target values.

- To predict the response $Y_{i,t+k}$ from the training set for the best

$$F_m(x_{i,t}) = F_{m-1}(x_{i,t}) + h_m(x_{i,t}) = Y_{i,t+k} \tag{3.6}$$

$$h_m(x_{i,t}) = Y_{i,t+k} - F_{m-1}(x_{i,t}) \tag{3.7}$$

- The new model $h_m(x)$ is chosen to be the negative gradient of the loss function concerning the current model's predictions:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \tag{3.8}$$

$$r_{m,i,t} = -\frac{\partial \frac{1}{2}\left(Y_{i,t+k} - F_{m-1}(x_{i,t})\right)^2}{\partial F_{m-1}(x_{i,t})} \tag{3.9}$$

- The model is updated by adding the new tree, weighted by a learning rate v.

$$F_m = F_{m-1}(x) + vh_m(x) = y \tag{3.10}$$

4. RESULTS AND DISCUSSION

Two models, Random Forest (RF) and Gradient Boosting Decision Trees (GBDT), were trained on the same dataset to predict hydroelectric power production. Their performances were evaluated and compared using R-squared (R2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) as key metrics across a range of tree counts (1-100 for RF, 1-100 for GBDT).

$$R^2(\%) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{4.1}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{4.2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4.3}$$

where y is actual or measured output, $\hat{y}$ is predicted output, $\overline{y}$ is the mean of y, and n indicates the number of observations (Zor et al., 2020).

Simple random sampling was performed to split 80% for training and 20% for testing.

RF model was constructed using 100 trees, each with a depth of 16 and varying numbers of leaves (1622-1885). This depth represents the maximum number of decisions a tree makes before reaching a prediction, while the number of leaves indicates the possible outcomes for a given input.

GBDT model was constructed using 100 trees, each with a depth of 6 and approximately 150.22 group splits. Each tree focuses on different features and split points to progressively enhance the model's predictive accuracy.

Table 4.1. Test Set Results

Model	Number of Trees	RMSE (MWh)	R^2 (%)	MAE (MWh)	Comp. Time (s)
RF	90	4.450	98.6	1.331	0.852
GBDT	100	5.449	97.9	2.351	0.256

Table 4.1 summarises the performance metrics for both RF and GBDT models. The results reveal that both models effectively forecast hydroelectric power production, albeit with subtle variations in accuracy and computational efficiency.

The RF model exhibited a higher R^2 value of 98.6%, surpassing the GBDT model's 97.9%. This suggests that the RF model explained a slightly greater proportion of the variance in hydroelectric power production, indicating a marginally better fit to the observed data.

Regarding prediction accuracy, the RF model consistently outperformed the GBDT model. The RF model's RMSE of 4.45 MWh was notably lower than GBDT's 5.449 MWh. This signifies that the RF model's predictions were, on average, closer to the actual hydroelectric power production values, resulting in a mean absolute percentage error improvement of 18.33% over GBDT. This substantial enhancement underscores the superior accuracy of the RF model in this context.

Furthermore, the RF model maintained its advantage in prediction accuracy with an MAE of 1.331 MWh, considerably lower than GBDT's 2.351 MWh. This further solidifies the conclusion that the RF model's predictions deviated less, on average, from the actual hydroelectric power production values.

However, the GBDT model demonstrated a significant advantage in computational efficiency. With a computation time of 0.256 seconds, the GBDT model outperformed the RF model by 233%, as the latter required 0.852 seconds to generate predictions. This remarkable difference highlights the GBDT model's exceptional computational speed in this forecasting task.

Table 4.2. Top 10 Results of RF

Model	Number of Trees	RMSE (MWh)	R2 (%)	MAE (MWh)	Comp. Time (s)
RF	90	4.450	98.6	1.331	0.852
RF	85	4.450	98.6	1.332	1.185
RF	91	4.451	98.6	1.331	0.927
RF	100	4.451	98.6	1.330	0.985
RF	87	4.452	98.6	1.332	0.852
RF	86	4.452	98.6	1.332	0.929
RF	99	4.452	98.6	1.330	0.999
RF	92	4.453	98.6	1.331	0.951
RF	97	4.453	98.6	1.331	0.986
RF	88	4.454	98.6	1.332	0.840

Table 4.3. Top 10 Results of GBDT

Model	Number of Trees	RMSE (MWh)	R2 (%)	MAE (MWh)	Comp. Time (s)
GBDT	100	5.449	97.9	2.351	0.256
GBDT	99	5.458	97.9	2.355	0.218
GBDT	98	5.465	97.9	2.357	0.228
GBDT	97	5.479	97.9	2.361	0.254
GBDT	96	5.491	97.9	2.366	0.364
GBDT	95	5.525	97.8	2.380	0.252
GBDT	94	5.537	97.8	2.384	0.205
GBDT	93	5.567	97.8	2.398	0.217
GBDT	92	5.572	97.8	2.400	0.218
GBDT	91	5.58	97.8	2.403	0.207

Tables 4.2 and 4.3 provide a more detailed breakdown of the top 10 performing models for both RF and GBDT, respectively. The results show that even with minor variations in hyperparameters, both models consistently achieve high levels of accuracy in predicting hydroelectric power production.

The performance of the RF model can be attributed to its ensemble nature. Multiple decision trees work together to improve predictive accuracy. The random selection of features for each split reduces the likelihood of overfitting, making the model robust and generalisable.

In contrast, the GBDT model, while a powerful tool, is sensitive to hyperparameters and requires careful tuning. The iterative boosting process, which corrects errors of previous trees, can sometimes lead to overfitting if not adequately managed. It is important to note these challenges and the need for parameter optimisation when using the GBDT model.

In conclusion, while the RF and GBDT models effectively forecast hydroelectric power production, the RF model exhibited superior predictive accuracy across multiple metrics.

However, the GBDT model compensated for its slightly lower accuracy with exceptional computational efficiency, making it over twice as fast as the RF model in generating predictions.

The findings of this study have significant practical implications. Accurate forecasting of hydroelectric power generation enables better grid management, optimising the balance between supply and demand. This is particularly important in regions with variable renewable energy sources, where predicting generation can reduce the reliance on fossil fuels and improve grid stability.

Furthermore, these models can assist in operational planning, allowing for better scheduling of maintenance activities and resource allocation. The improved accuracy of the RF model provides a reliable tool for energy planners and operators to make informed decisions, enhancing the efficiency and sustainability of HPPs.

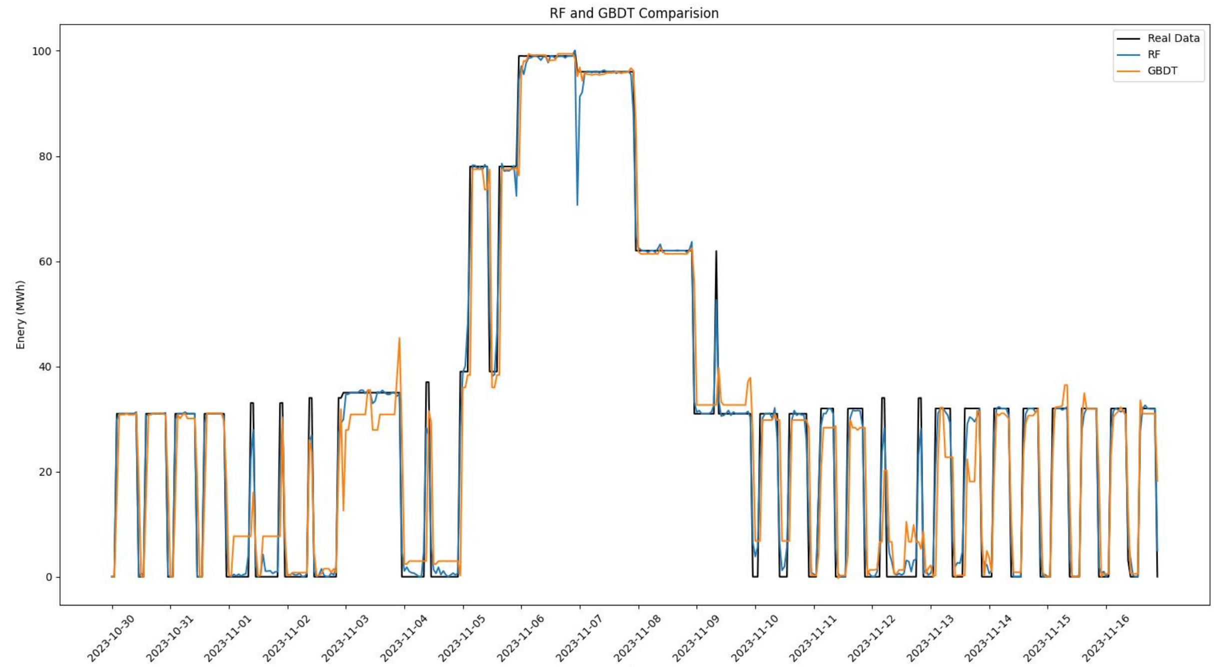

Figure 4.1. Compared data

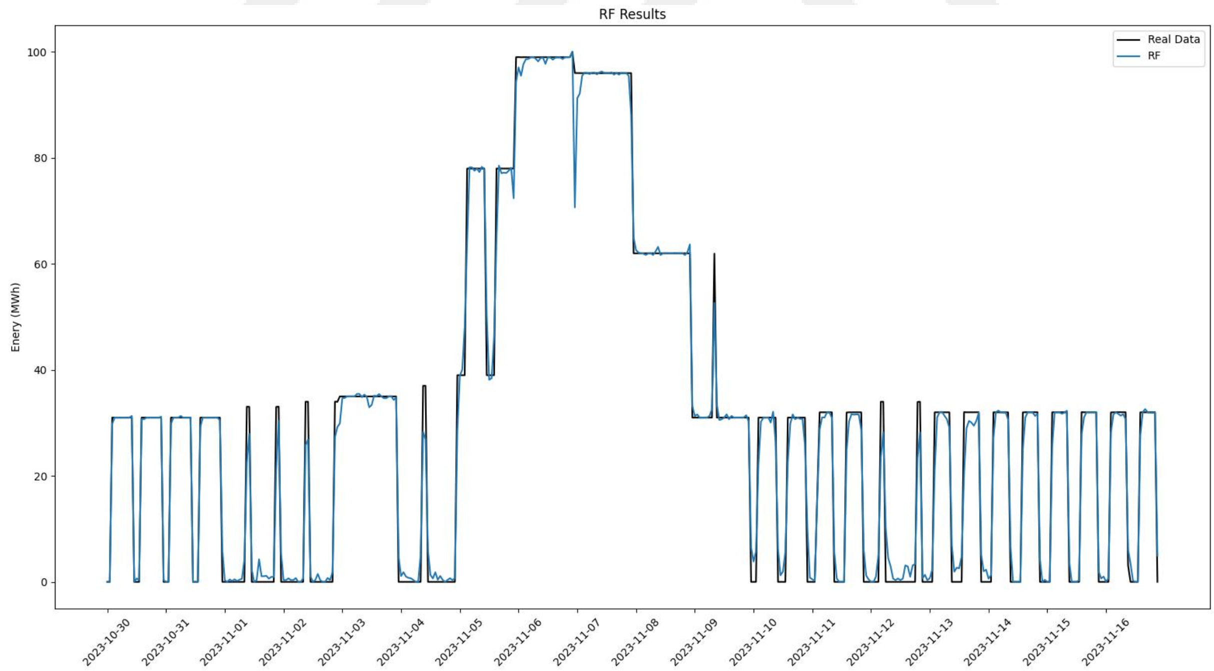

Figure 4.2. RF Results

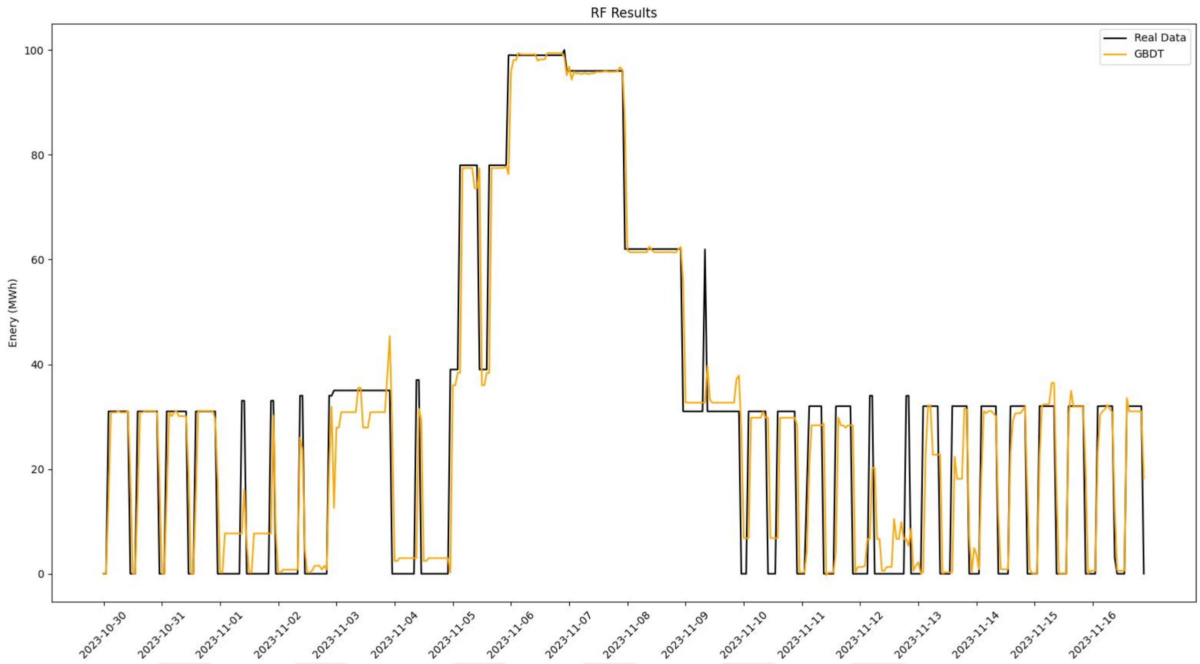

Figure 4.3. GBDT Results

The choice of RF and GBDT depends on the goals and the specific application. GBDT might be preferable for real-time dam operations due to its inherent parallelism, allowing for faster predictions when analysing incoming data. However, RF might be favoured for long-term planning efforts that prioritise accuracy.

The level of interpretability required in a hydroelectric power system can also influence model choice. With its ensemble of decision trees, RF can provide some level of interpretability through feature importance scores. However, GBDT can be more challenging to interpret, especially with deeper trees and complex feature interactions.

Lastly, model maintenance is another crucial consideration. RF models can be relatively stable over time, as the randomness inherent in their construction provides some robustness to changes in the underlying data. GBDT models, however, might require more frequent retraining as they can be more sensitive to changes in the data distribution.

5. CONCLUSIONS

The study has shown promising results in forecasting hydroelectric power production at the Aslantaş HPP by utilising historical power generation, meteorological, and calendar data with tree-based ML models, namely RF and GBDT. Proposed models have demonstrated reliable performance in predicting future hydroelectric power generation.

The RF and GBDT models have consistently demonstrated robust performance in forecasting hydroelectric power production. This reliability paves the way for more informed decision-making in resource allocation, reservoir management, and grid integration, all crucial aspects for a sustainable future. The RF model shows promising results in enhancing the precision of hydroelectric power forecasting. Its contribution to sustainable energy practices is significant, ensuring the efficient use of natural resources for electricity generation and water management. These results should inspire confidence in the effectiveness of these models for our professional colleagues and researchers in the field.

The research advances the understanding of hydroelectric power forecasting and offers practical opportunities for decision-makers in the energy sector. Incorporating AI-based algorithms into forecasting models is essential to meet the increasing demand for renewable energy. Future research may explore more AI-based algorithms and optimisation techniques to enhance accuracy in diverse operational settings.

Accurate forecasting of hydroelectric power generation is crucial for efficient grid management, optimising the balance between supply and demand. The RF model's superior performance provides a reliable tool for energy planners and operators, enabling better scheduling of maintenance activities, resource allocation, and operational planning. This leads to improved efficiency and sustainability of HPPs. Precise hydroelectric power forecasting reduces reliance on fossil fuels, lowering carbon emissions and promoting environmental sustainability. Additionally, it supports the economic planning of electricity generation, optimising investment strategies for hydroelectric projects, and ensuring the stable and efficient operation of the power grid.

6. RECOMMENDATIONS

Future research should explore other AI-based algorithms, such as deep learning models (e.g., LSTM, CNN), hybrid models, and ensemble methods, to further enhance the accuracy and robustness of hydroelectric power forecasting.

Incorporating real-time data from various sources, such as satellite imagery, advanced meteorological forecasts, and IoT-based sensors, can improve the timeliness and accuracy of predictions. This real-time integration will help address the dynamic nature of hydroelectric power generation.

It is important to note that these results are specific to the dataset and parameters used in this study. The performance of the models may vary depending on the data's characteristics and the models' specific configuration. Further research is needed to explore the generalizability of these findings to other datasets and forecasting scenarios.

Implementing advanced optimisation techniques, such as genetic algorithms, particle swarm optimisation, and grid search, can fine-tune model parameters, enhancing the predictive performance of AI models.

Extending the research to different types of HPPs, including run-of-river, pumped storage, and small-scale hydro plants, will validate the generalizability of the proposed models and identify any unique challenges associated with these settings.

Developing models for long-term forecasting can aid in strategic planning and policymaking, ensuring sustainable energy development and resilience against climate change impacts.

The study advances the understanding of hydroelectric power forecasting using AI-based algorithms, offering practical tools for enhancing the efficiency and sustainability of HPPs. The RF model shows excellent potential for improving the precision of hydroelectric power forecasts, supporting better decision-making in resource management and grid operations. By incorporating AI into hydroelectric power forecasting, the energy sector can move towards more sustainable and reliable energy solutions, contributing to global efforts in combating climate change and promoting renewable energy adoption.

REFERENCES

Abu, N., Tukimat, N. N. A., & Abu, N. (2024). *Forecasting of hydropower production using Box-Jenkins model at Tasik Kenyir, Terengganu*. 050005. https://doi.org/10.1063/5.0193157

Acaroğlu, H., Kartal, H. M., & García Márquez, F. P. (2023). Testing the environmental Kuznets curve hypothesis in terms of ecological footprint and CO2 emissions through energy diversification for Turkey. *Environmental Science and Pollution Research*, *30*(22), 63289–63304. https://doi.org/10.1007/S11356-023-26278-W/FIGURES/3

Achig, R. S., Gonzales, J. A., & Hidalgo, C. A. (2023). *Flow Prediction for Hydropower Generation Using LMST Neural Networks*. https://doi.org/10.1088/1742-6596/2609/1/012006

Afonso De Sousa, M. (2019). *Development of Electricity Generation Forecast Models Applied to Small Hydroelectric Plants* [Master Thesis, Pontifical Catholic University of Rio de Janeiro]. https://doi.org/10.17771/PUCRio.acad.47236

Ahmad, S. K., & Hossain, F. (2020). Maximizing energy production from hydropower dams using short-term weather forecasts. *Renewable Energy*, *146*, 1560–1577. https://doi.org/10.1016/j.renene.2019.07.126

Aiken, E. (2018). *Script to download data from MERRA-2 database based on lat/lon coordinates*. Github. https://github.com/emilylaiken/merradownload

Al Rayess, H. M., & Ülke Keskin, A. (2021). Forecasting the hydroelectric power generation of GCMs using machine learning techniques and deep learning (Almus Dam, Turkey). *Geofizika*, *38*(1), 1–14. https://doi.org/10.15233/gfz.2021.38.4

Aleksandrovskii, A. Yu., & Borshch, P. S. (2013). Prediction of electric-power generation at hydroelectric power plants. *Power Technology and Engineering*, *47*(2), 83–88. https://doi.org/10.1007/s10749-013-0403-8

Barzola-Monteses, J., Gómez-Romero, J., Espinoza-Andaluz, M., & Fajardo, W. (2022). Hydropower production prediction using artificial neural networks: an Ecuadorian application case. *Neural Computing and Applications*, *34*(16), 13253–13266. https://doi.org/10.1007/s00521-021-06746-5

Barzola-Monteses, J., Mite-León, M., Espinoza-Andaluz, M., Gómez-Romero, J., & Fajardo, W. (2019). Time series analysis for predicting hydroelectric power production: The ecuador case. *Sustainability (Switzerland)*, *11*(23). https://doi.org/10.3390/SU11236539

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/S11749-016-0481-7/FIGURES/4

Bilgili, M., Keiyinci, S., & Ekinci, F. (2022). One-day ahead forecasting of energy production from run-of-river hydroelectric power plants with a deep learning approach. *Scientia Iranica*, *0*(0), 0–0. https://doi.org/10.24200/sci.2022.58636.5825

Breiman, L. (2001). *Random Forests*. *45*, 5–32.

Çakır, S. (2023). Renewable energy generation forecasting in Turkey via intuitionistic fuzzy time series approach. *Renewable Energy*, *214*, 194–200. https://doi.org/10.1016/j.renene.2023.05.132

Cassagnole, M., Ramos, M.-H., Zalachori, I., Thirel, G., Garçon, R., Gailhard, J., & Ouillon, T. (2021). Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs – a conceptual approach. *Hydrology and Earth System Sciences*, *25*(2), 1033–1052. https://doi.org/10.5194/hess-25-1033-2021

Cheganova, N. F., Matrenin, P. V., & Mitrofanov, S. V. (2023). Medium-Term Forecast of Water Inflow to the Sayano-Shushenskaya Hydroelectric Power Plant Reservoir by Autoregressive Machine Learning Models. *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE)*, 470–473. https://doi.org/10.1109/APEIE59731.2023.10347751

Coadou, Y. (2022). Boosted Decision Trees. In *Artificial Intelligence for High Energy Physics* (pp. 9–58). WORLD SCIENTIFIC. https://doi.org/10.1142/9789811234033_0002

Corbari, C., Ravazzani, G., Perotto, A., Lanzingher, G., Lombardi, G., Quadrio, M., Mancini, M., & Salerno, R. (2022). Weekly Monitoring and Forecasting of Hydropower Production Coupling Meteo-Hydrological Modeling with Ground and Satellite Data in the Italian Alps. *Hydrology*, *9*(2). https://doi.org/10.3390/hydrology9020029

de Sousa, M. A., Maçaira, P. M., Souza, R. C., Cyrino Oliveira, F. L., & Calili, R. F. (2020). *Forecasting Electricity Generation of Small Hydropower Plants* (pp. 45–54). https://doi.org/10.1007/978-3-030-23816-2_5

Dehghani, M., Riahi-Madvar, H., Hooshyaripor, F., Mosavi, A., Shamshirband, S., Zavadskas, E., & Chau, K. (2019). Prediction of Hydropower Generation Using Grey Wolf Optimization Adaptive Neuro-Fuzzy Inference System. *Energies*, *12*(2), 289. https://doi.org/10.3390/en12020289

Demir, M. (2015). *By Using Artificial Neural Networks to Determine The Ouantity Of Production Energy: Example Of Dim Dam* [M.Sc.]. Süleyman Demirel University, Graduate School of Applied and Natural Sciences.

Deng, H., Zhou, Y., Wang, L., & Zhang, C. (2021). Ensemble learning for the early prediction of neonatal jaundice with genetic features. *BMC Medical Informatics and Decision Making*, *21*(1), 1–11. https://doi.org/10.1186/S12911-021-01701-9/FIGURES/6

Dmitrieva, K. (2015). *Forecasting of a hydropower plant energy production*. Ostfold Universitt Collage.

Drakaki, K.-K., Drakaki, K.-K., Sakki, G.-K., Sakki, G.-K., Tsoukalas, I., Tsoukalas, I., Kossieris, P., Kossieris, P., Efstratiadis, A., & Efstratiadis, A. (2022). Day-ahead energy production in small hydropower plants: uncertainty-aware forecasts through effective coupling of knowledge and data. *Advances in Geosciences*. https://doi.org/10.5194/adgeo-56-155-2022

Ehtearm, M., Ghayoumi Zadeh, H., Seifi, A., Fayazi, A., & Dehghani, M. (2023). Predicting Hydropower Production Using Deep Learning CNN-ANN Hybridized with Gaussian Process Regression and Salp Algorithm. *Water Resources Management*, *37*(9), 3671–3697. https://doi.org/10.1007/s11269-023-03521-0

Ekanayake, P., Wickramasinghe, L., Jayasinghe, J. M. J. W., & Rathnayake, U. (2021). Regression-Based Prediction of Power Generation at Samanalawewa Hydropower Plant in Sri Lanka Using Machine Learning. *Mathematical Problems in Engineering*, *2021*, 1–12. https://doi.org/10.1155/2021/4913824

Energy Info Bank of Turkey. (n.d.). Republic of Türkiye Ministry of Energy and Natural Resources. Retrieved 30 April 2024, from https://enerji.gov.tr/infobank-energy-electricity

EÜAŞ Aslantaş HPP web page. (n.d.). EÜAŞ. Retrieved 18 April 2024, from https://www.euas.gov.tr/en-US/power-plants/aslantas-HPP

Fan Qiang, Lin Chenghui, Wen Xiankui, Peng, W., & Chen Yuanyuan. (2016). Research on application of run-of-river small Hydropower Station group short-term power forecast system in Guizhou Power Grid. *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, 328–334. https://doi.org/10.1109/ICPRE.2016.7871226

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Global Modeling and Assimilation Office (GMAO) (2015). (n.d.-a). *MERRA-2 inst1_2d_asm_Nx: 2d,1-Hourly,Instantaneous,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4 (M2I1NXASM)*. Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC). https://doi.org/10.5067/3Z173KIE2TPD

Global Modeling and Assimilation Office (GMAO) (2015). (n.d.-b). *MERRA-2 tavg1_2d_rad_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Radiation*

Diagnostics V5.12.4 (M2T1NXRAD). Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/Q9QMY5PBNV1T

Hammid, A. T., Sulaiman, M. H. Bin, & Abdalla, A. N. (2018). Prediction of small hydropower plant power production in Himreen Lake dam (HLD) using artificial neural network. *Alexandria Engineering Journal*, *57*(1), 211–221. https://doi.org/10.1016/j.aej.2016.12.011

Huangpeng, Q., Huang, W., & Gholinia, F. (2021). Forecast of the hydropower generation under influence of climate change based on RCPs and Developed Crow Search Optimization Algorithm. *Energy Reports*, *7*, 385–397. https://doi.org/10.1016/j.egyr.2021.01.006

Hussin, S. N. H. S., Malek, M. A., Jaddi, N. S., & Hamid, Z. A. (2016). Hybrid metaheuristic of artificial neural network — Bat algorithm in forecasting electricity production and water consumption at Sultan Azlan shah Hydropower plant. *2016 IEEE International Conference on Power and Energy (PECon)*, 28–31. https://doi.org/10.1109/PECON.2016.7951467

İNAL, S., AKKAYA OY, S., & ÖZDEMİR, A. E. (2023). A Neural Network Model for Estimation of Maximum Next Day Energy Generation Capacity of a Hydropower Station: A Case Study from Turkey. *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*, *19*(3), 197–204. https://doi.org/10.18466/cbayarfbe.1218381

Javed, U., Fraz, M. M., Mahmood, I., Shahzad, M., & Arif, O. (2020). Forecasting of Electricity Generation for Hydro Power Plants. *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, 32–36. https://doi.org/10.1109/HONET50430.2020.9322841

Ji, Q., Zhang, S., Duan, Q., Gong, Y., Li, Y., Xie, X., Bai, J., Huang, C., & Zhao, X. (2022). Short- and Medium-Term Power Demand Forecasting with Multiple Factors Based on Multi-Model Fusion. *Mathematics*, *10*(12), 2148. https://doi.org/10.3390/math10122148

Joaquim, P. C. E., Joaquim, P. C. E., Rosa, J. L. G., & Rosa, J. L. G. (2005). *Artificial neural networks for temporal processing applied to prediction of electric energy in small hydroelectric power stations*. https://doi.org/10.1109/ijcnn.2005.1556317

Kiruthiga, D., & Amudha, T. (2019). *Hydropower Generation Optimization and Forecasting Using PSO* (pp. 411–421). https://doi.org/10.1007/978-981-10-8055-5_37

Konica, J. A., & Staka, E. (2017). Forecasting of a hydropower plant energy production with Fuzzy logic Case for Albania. *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, *4*, 2458–9403. www.jmest.orgJMESTN423521827244

Kumar, K., Saini, G., Kumar, N., Kaiser, M. S., Kannan, R., & Shah, R. (2022). Prediction of energy generation target of hydropower plants using artificial neural networks. In *Sustainable Developments by Artificial Intelligence and Machine Learning for Renewable Energies* (pp. 309–320). Elsevier. https://doi.org/10.1016/B978-0-323-91228-0.00005-7

Li, G., Liu, C.-X., Liao, S.-L., & Cheng, C.-T. (2015). Applying a Correlation Analysis Method to Long-Term Forecasting of Power Production at Small Hydropower Plants. *Water*, *7*(12), 4806–4820. https://doi.org/10.3390/w7094806

Li, G., Sun, Y., He, Y., Li, X., & Tu, Q. (2014). Short-Term Power Generation Energy Forecasting Model for Small Hydropower Stations Using GA-SVM. *Mathematical Problems in Engineering*, *2014*, 1–9. https://doi.org/10.1155/2014/381387

Li, G.-D., Masuda, S., & Nagai, M. (2016). Prediction of hydroelectric power generation in Japan. *Energy Sources, Part B: Economics, Planning, and Policy*, *11*(3), 288–294. https://doi.org/10.1080/15567249.2012.708097

Li, L., Yao, F., Huang, Y., & Zhou, F. (2019). Hydropower Generation Forecasting via Deep Neural Network. *2019 6th International Conference on Information Science and Control Engineering (ICISCE)*, 324–328. https://doi.org/10.1109/ICISCE48695.2019.00071

Li, M., Deng, C. H., Tan, J., Yang, W., & Zheng, L. (2016). Research on Small Hydropower Generation Forecasting Method Based on Improved BP Neural Network. *Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control*. https://doi.org/10.2991/icmemtc-16.2016.214

Li, Z., Hu, X., Guo, H., & Xiong, X. (2023). A novel Weighted Average Weakening Buffer Operator based Fractional order accumulation Seasonal Grouping Grey Model for predicting the hydropower generation. *Energy*, *277*, 127568. https://doi.org/10.1016/j.energy.2023.127568

Lian, C., Lian, C., Lian, C., Wang, H., Wei, H., Wei, H., Qin, S., Qin, S., Li, Z., & Li, Z. (2020). *Trend-guided Small Hydropower System Power Prediction Based on Extreme Learning Machine*. https://doi.org/10.1109/icpee51316.2020.9311091

Lopes, M. N. G., da Rocha, B. R. P., Vieira, A. C., de Sá, J. A. S., Rolim, P. A. M., & da Silva, A. G. (2019). Artificial neural networks approaches for predicting the potential for hydropower generation: a case study for Amazon region. *Journal of Intelligent & Fuzzy Systems*, *36*(6), 5757–5772. https://doi.org/10.3233/JIFS-181604

Luwe, Y. J., Lee, C. P., & Lim, K. M. (2023). Wearable sensor-based human activity recognition with ensemble learning: a comparison study. *International Journal of Electrical and Computer Engineering*, *13*(4), 4029–4040. https://doi.org/10.11591/IJECE.V13I4.PP4029-4040

Mnguu, E. E. (2020). *FORECASTING HYDRO-POWER GENERATION USING ARIMA MODEL: A CASE OF MTERA DAM IN TANZANIA* [Master Thesis]. The University Of Dodoma.

Monteiro, C., Ramirez-Rosado, I. J., & Fernandez-Jimenez, L. A. (2013). Short-term forecasting model for electric power production of small-hydro power plants. *Renewable Energy*, *50*, 387–394. https://doi.org/10.1016/j.renene.2012.06.061

Monteiro, C., Ramirez-Rosado, I. J., & Fernandez-Jimenez, L. A. (2014). Short-term forecasting model for aggregated regional hydropower generation. *Energy Conversion and Management*, *88*, 231–238. https://doi.org/10.1016/j.enconman.2014.08.017

Polprasert, J., Hanh Nguyen, V. A., & Nathanael Charoensook, S. (2021). Forecasting Models for Hydropower Production Using ARIMA Method. *2021 9th International Electrical Engineering Congress (IEECON)*, 197–200. https://doi.org/10.1109/iEECON51072.2021.9440293

Prakash, S. A., Shah, D., Jayavel, K., & Mtonga, K. (2022). Hydropower Energy Generation Prediction Model: A Machine Learning Approch. *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 01–04. https://doi.org/10.1109/ICCCI54379.2022.9740915

Rahman, M. M., Shakeri, M., Tiong, S. K., Khatun, F., Amin, N., Pasupuleti, J., & Hasan, M. K. (2021). Prospective Methodologies in Hybrid Renewable Energy Systems for Energy Prediction Using Artificial Neural Networks. *Sustainability*, *13*(4), 2393. https://doi.org/10.3390/su13042393

Rathnayake, N., Rathnayake, U., Dang, T. L., & Hoshino, Y. (2022). A Cascaded Adaptive Network-Based Fuzzy Inference System for Hydropower Forecasting. *Sensors*, *22*(8). https://doi.org/10.3390/s22082905

Razi, M., Yusuff, M. A., Tee, B. T., & Zakaria, K. A. (2017). Prediction of available power being generate in small hydropower system at Sungai Perting Bentong Pahang. *MATEC Web of Conferences*, *90*, 01028. https://doi.org/10.1051/matecconf/20179001028

Ren, X., Zhao, Y., Hao, D., Sun, Y., Chen, S., & Gholinia, F. (2021). Predicting optimal hydropower generation with help optimal management of water resources by Developed Wildebeest Herd Optimization (DWHO). *Energy Reports*, *7*, 968–980. https://doi.org/10.1016/j.egyr.2021.02.007

Sarpong, S. A., & Agyei, A. (2022). Forecasting Hydropower Generation in Ghana Using ARIMA Models. *International Journal of Statistics and Probability*, *11*(5), 30. https://doi.org/10.5539/ijsp.v11n5p30

Sattar Hanoon, M., Najah Ahmed, A., Razzaq, A., Oudah, A. Y., Alkhayyat, A., Feng Huang, Y., kumar, P., & El-Shafie, A. (2023). Prediction of hydropower generation via machine learning algorithms at three Gorges Dam, China. *Ain Shams Engineering Journal*, *14*(4), 101919. https://doi.org/10.1016/j.asej.2022.101919

Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests. *Https://Doi.Org/10.1214/15-AOS1321*, *43*(4), 1716–1741. https://doi.org/10.1214/15-AOS1321

Sessa, V., Assoumou, E., Bossy, M., & Simões, S. G. (2021). Analyzing the Applicability of Random Forest-Based Models for the Forecast of Run-of-River Hydropower Generation. *Clean Technologies*, *3*(4), 858–880. https://doi.org/10.3390/cleantechnol3040050

Tebong, N. K., Simo, T., & Takougang, A. N. (2023). Two-level deep learning ensemble model for forecasting hydroelectricity production. *Energy Reports*, *10*, 2793–2803. https://doi.org/10.1016/J.EGYR.2023.09.091

the EPIAS Transparency Platform. (n.d.). *Real-Time Generation Data | the EPIAS Transparency Platform*. Retrieved 19 May 2024, from https://seffaflik.epias.com.tr/electricity/electricity-generation/ex-post-generation/real-time-generation

TÜBİTAK MAM. (n.d.). *Project ATHOM - Flow Forecasting And Catchment Optimisation Model*. Retrieved 19 March 2024, from https://mam.tubitak.gov.tr/tr/haber/tubitak-mam-akim-tahmin-ve-havza-optimizasyon-modeli-athom-projesi-genel-degerlendirme

Uzlu, E., Akpınar, A., Özturk, H. T., Nacar, S., & Kankal, M. (2014). Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey. *Energy*, *69*, 638–647. https://doi.org/10.1016/j.energy.2014.03.059

Velasquez, V., & Flores, W. (2022). Machine Learning Approach for Predictive Maintenance in Hydroelectric Power Plants. *2022 IEEE Biennial Congress of Argentina (ARGENCON)*, 1–6. https://doi.org/10.1109/ARGENCON55245.2022.9939782

Wang, J., Gao, Z., & Ma, Y. (2022). Prediction Model of Hydropower Generation and Its Economic Benefits Based on EEMD-ADAM-GRU Fusion Model. *Water*, *14*(23), 3896. https://doi.org/10.3390/w14233896

Wang, S., Wang, S., Wang, S., Yu, L., Yu, L., Tang, L., Tang, L., Tang, L., Wang, S., Wang, S., & Wang, S. (2011). A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. *Energy*. https://doi.org/10.1016/j.energy.2011.09.010

Wang, Z.-X., Li, Q., & Pei, L.-L. (2017). Grey forecasting method of quarterly hydropower production in China based on a data grouping approach. *Applied Mathematical Modelling*, *51*, 302–316. https://doi.org/10.1016/j.apm.2017.07.003

Wei, Z., Shen, X., Qiu, G., Liu, Y., & Liu, J. (2024). An Integrated Transfer Learning Method for Power Generation Prediction of Run-Off Small Hydropower in Data-Scarce Areas. *IEEE Transactions on Smart Grid*, *15*(1), 1030–1041. https://doi.org/10.1109/TSG.2023.3276390

Wergeland, K. K. (2023). *Predicting next day's production in run-of-river hydropower plants using machine learning* [Master Thesis]. University of Bergen.

Xia, D., Luo, X., Bao, X., Liu, Y., & Lin, Z. (2022). Study on Forecasting Model of Power Generation of Small Hydropower Station. *10th Academic Conference of Geology Resource Management and Sustainable Development 2022*, 757–770. https://doi.org/10.52202/067798-0101

Yang, S., Wei, H., Zhang, L., & Qin, S. (2021). Daily Power Generation Forecasting Method for a Group of Small Hydropower Stations Considering the Spatial and Temporal Distribution of Precipitation—South China Case Study. *Energies*, *14*(15), 4387. https://doi.org/10.3390/en14154387

Yang, Z., Yang, Z., Wang, Y., Wang, Y., Yang, K., & Yang, K. (2021). The stochastic short-term hydropower generation scheduling considering uncertainty in load output forecasts. *Energy*. https://doi.org/10.1016/j.energy.2021.122838

Yildiz, C., & Açikgöz, H. (2021). Forecasting diversion type hydropower plant generations using an artificial bee colony based extreme learning machine method. *Energy Sources, Part B: Economics, Planning, and Policy*, *16*(2), 216–234. https://doi.org/10.1080/15567249.2021.1872119

Zeng, B., He, C., Mao, C., & Wu, Y. (2023). Forecasting China's hydropower generation capacity using a novel grey combination optimization model. *Energy*, *262*, 125341. https://doi.org/10.1016/j.energy.2022.125341

Zhang, H., Shen, X., Wang, Y., Wang, J., & Li, B. (2023). Generation Prediction for Hydropower Considering the Coupling Effect of Adjacent Plants. *2023 Panda Forum on Power and Energy (PandaFPE)*, 1967–1971. https://doi.org/10.1109/PandaFPE57779.2023.10140847

Zhang, S., Chen, S.-J., Ma, G., Huang, W., & Li, B. (2022). Generation hybrid forecasting for frequency-modulation hydropower station based on improved EEMD and ANN adaptive switching. *Electrical Engineering*, *104*(5), 2949–2966. https://doi.org/10.1007/s00202-022-01526-3

Zhang, Z., & Jung, C. (2019). GBDT-MO: Gradient Boosted Decision Trees for Multiple Outputs. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(7), 3156–3167. https://doi.org/10.1109/TNNLS.2020.3009776

Zhong, W., Guo, J., Chen, L., Zhou, J., Zhang, J., & Wang, D. (2020). Future hydropower generation prediction of large-scale reservoirs in the upper Yangtze River basin under climate change. *Journal of Hydrology*, *588*, 125013. https://doi.org/10.1016/j.jhydrol.2020.125013

Zhou, F., Li, L., Zhang, K., Trajcevski, G., Yao, F., Huang, Y., Zhong, T., Wang, J., & Liu, Q. (2020). Forecasting the Evolution of Hydropower Generation. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2861–2870. https://doi.org/10.1145/3394486.3403337

Zolfaghari, M., & Golabi, M. R. (2021). Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models. *Renewable Energy*, *170*, 1367–1381. https://doi.org/10.1016/j.renene.2021.02.017

Zor, K., & Çelik, Ö. (2021). Short-Term Nonindustrial Reactive Power Forecasting. *17th International Conference of Young Scientists on Energy and Natural Sciences Issues (CYSENI2021)*.

Zor, K., Çelik, Ö., Timur, O., & Teke, A. (2020). Short-Term Building Electrical Energy Consumption Forecasting by Employing Gene Expression Programming and GMDH Networks. *Energies*, *13*(5), 1102. https://doi.org/10.3390/en13051102

Zou, Y., Chen, H., Zhang, Y., Wang, Z., Pang, L., Lan, X., Zhao, Y., Liu, T., Wang, B., & Peng, R. (2023). Multimode hydropower power prediction based on long short-term memory. In S. Jia & H. Dong (Eds.), *Sixth International Conference on Computer Information Science and Application Technology (CISAT 2023)* (p. 233). SPIE. https://doi.org/10.1117/12.3004172