ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

Deep Learning Models for Future Hydropower Production Prediction: A Case Study of the Koka Hydroelectric Power Plant, Ethiopia

by

Ababa Lata Guta

Advisor: Hailemichael Kebede (PHD)

This Thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science

in

Computational Data Science(Data Science)

Addis Ababa

Ethiopia

June, 2024

ADDIS ABABA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE
COMPUTATIONAL DATA SCIENCE PROGRAM

This is to certify that the thesis prepared by Ababa Lata, entitled: Deep Learning Models for Future Hydropower Production Prediction: A Case Study of the Koka Hydroelectric Power Plant, Ethiopia, and submitted in partial fulfillment of the requirements for the degree of Master of Science in Computational Data Science in specialization in Data Science complies with the regulations of the university and meets the accepted standards with respect to originality and quality.
Signed by the examining committee:

	Sign	Date
Examiner ____________________	__________	__________
(External examiner)		
Examiner ____________________	__________	__________
(Internal Examiner)		
Advisor ____________________	__________	__________
(Hailemichael Kebede (PHD))		
Co-Advisor ____________________	__________	__________
(Behailu Birhanu (PHD))		

Abstract

Hydropower, a clean and renewable energy source driven by the water cycle, plays a crucial role in many countries. Predicting future hydropower production is vital for strategic decision-making and optimizing energy resource utilization. This study evaluated three deep learning models (Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gate Recurrent Unit (GRU)) for daily and weekly hydropower energy production future prediction at the Koka Dam in Ethiopia's Awash basin. The models aimed to forecast not only daily and weekly production but also the next year of generation. Daily and weekly hydropower production, precipitation, minimum temperature, maximum temperature, relative humidity, maximum wind speed, minimum wind speed, wind direction, and all sky-surface short wave data from September 2010 to November 2023 were used. After preprocessing, the data was split into training and testing sets for model training and evaluation respectively. Performance metrics like R-squared, MAE, MSE, and RMSE were calculated for each model. The GRU model emerged as the best performer, achieving an R-squared value of 0.9920, MAE of 8.6121, MSE of 143.5549, and RMSE of 11.9814 for daily hydropower energy generation prediction, and R-squared of 0.9960, MAE of 5.9925, MSE of 66.2187, and RMSE of 8.1375 for weekly hydropower energy generation prediction. This superior model was then employed to predict the Koka Dam's daily and weekly hydropower energy production for the next 1 year. The study identified the right hydro-power production prediction model for future potential. Furthermore, this outcome can help to maximize the use of medium hydro-power resources, contributing to the region's energy security and sustainable development.
Keywords: Hydro-power, GRU, LSTM, RNN, Metrics, Prediction

Acknowledgment

I am filled with immense gratitude as I reflect on the successful completion of this significant project. First and foremost, I am deeply thankful for the divine power that has guided and supported me throughout this journey. Next, I would like to extend my heartfelt appreciation to express my sincere gratitude to my advisor, Doctor Hailemichael Kebede, and Doctor Behailu Birhanu, for their unwavering mentorship and guidance throughout the duration of this thesis project. Their expertise, constructive feedback, and unwavering encouragement have been pivotal in shaping the direction and quality of my work. Their dedication to nurturing the academic and professional growth of their students is truly commendable. I would also like to thank the Department of Computational Data Science program for providing me with this invaluable learning opportunity. The knowledge, resources, and support I have received from this esteemed academic institution have been truly transformative, allowing me to expand my understanding and skills in the field of data science.

Contents

List of Figures

List of Tables

List of Abbreviation

Symbol	Description
RNN	Recurrent Neural Network
$LSTM$	Long Short Term Memory
GRU	Gated Recurrent Unit
TWH	Tera Wat hour
GW	Giga wat
MW	Mega wat
DL	Deep Leaarning
ML	Machine Learning
$ARIMA$	AutoRegressive Integrated Moving Average
$SARIMA$	Seasonal AutoRegressive Integrated Moving Average
$EEMD$	Ensemble Empirical Mode Decomposition
$ADAM$	Adaptive Moment Estimation
SVM	Supportive Vector Machine
CNN	Convolutional Neural Network
ANN	Artificial neural network
$T2M-MAX$	Maximum Temperature
$RH2M$	Relative Humidity
ADF	Adfuller test
MAE	Mean Absolute Error
MSE	Mean Square Error
$RMSE$	Root Mean Square Error
R^2	R-Squared

1 General Introduction

1.1 Background

The development of a country depends on its work centers, the generation of business, and the advancement of its companies and administrations. Electricity is vital for each of these exercises. It is basic to extend energy generation and meet the growing demand. At the same time, it is fundamental to utilize energy resources that offer the best opportunity for environmental protection and guarantee the move to a worldwide economy based on renewable energy[13]. With this idea, Hydropower potential comes as one of the most relevant alternatives given the fact that it could be a non-pollutant resource, and it can be extracted with minimal impact on the surrounding environment. Besides, hydroelectric(hydropower) is a renewable energy source where electrical power is derived from the energy of water moving from higher to lower elevations. It is a proven, mature, predictable, and price-competitive technology. Nowadays Hydroelectricity generation expanded by nearly 70 TWh (up nearly to 2 %) in 2022, coming to 4300TWh. Hydropower remains the biggest renewable source of power, producing more than all other renewable advances combined. Within the Net Zero Outflows by 2050 Scenario, hydropower keeps up an average yearly era development rate of nearly 4 % in 2023-2030 to supply around 5500 TWh of power per year. Within the final five years, the normal development rate was less than one-third of what is required, flagging a requirement for essentially more grounded endeavors, particularly to streamline allowing and guarantee extended maintainability. Hydropower plants ought to be perceived as a dependable spine of clean power systems in the long run and supported in like manner[11][22][15].

These studies underscore the importance of energy generation, especially hydropower as a sustainable source, for national development. However, the current pace of hydro-power development falls short of requirements. To solidify hydropower's role as the backbone of clean energy, stronger efforts in permitting and sustainability are crucial.

Ethiopia, like many African nations, has underutilized energy resources. Biomass fuels dominate the energy supply, accounting for 86.5 % of the total 1,878,972 TJ. Oil follows at 10 %, with hydropower lagging at 2.8 % Solar and coal contribute minimally [20, 31].[18][29]. Based on division vitality utilization in Ethiopia's family segment devours around 88.2 percent of add up to vitality supply, taken after by the transportation segment (8.4 percent) and 3 % by industry[18][19]. Be that as it may, the lion's share of vitality supply in the family division subsequently is secured by bio-energy whereas a transport division is a transcendence run by imported petroleum[18][27]. In 2020, to power era capacity of the nation was 4,713 MW with 86 percent delivered from hydro, 6 percent from bio-energy, 7 percent from wind, and 1 percent from non-renewable sources. According to the IEA report in 2019, the country's national electrification rate is still very low, at around 48.3 percent[8]. Besides, there's a stark contrast in the rate of power get to between urban and rural areas; in urban zones, 92.8 percent of the populace has got to power, while in country zones, access to electricity remains greatly low at around 36.3 percent[20]. As a result, the endless larger part of the populace lives in rural areas and relies heavily on conventional biomass vitality sources for cooking and heating [29][19]. The assessed power utilization per capita of the country by 2017 was 100 kWh and expanded to approximately 130 kWh. In any case, this capacity is certainly too small considering the vitality resource potential of the nation for both conventional and renewable vitality utilization [18].

Hydropower can serve as a bridge into a transition to more sustainable sources of energy due to its large potential development scale, environmentally friendly nature, and the lower average cost of electricity generated than any other energy generation technology[18]. The technical hydropower potential in Africa is assessed to be around 283 GW and capable of producing near to 1200 TWH per year, bookkeeping for 8 % of the worldwide specialized potential. It is assessed that 90 % of the hydropower potential within the landmass is still undiscovered[20][17]. In 2018, about half of Africans (600 million individuals) needed to get to power, and generally 80 % of sub-Saharan African businesses experienced visit control blackouts, resulting in financial misfortunes. Moreover, more than 70 % of the population, or around 900 million individuals, need to get clean cooking fuel [18].

Ethiopia, one of East Africa's countries, is well-known for its inexhaustible water resources. There are twelve river basins within the country. The overall yearly cruel surface runoff from all twelve river basins is assessed to be 124 billion cubes. Eight of Ethiopia's twelve river basins have been distinguished for their hydroelectric power generation potential. Around 300 hydropower plant destinations have been distinguished in those eight river basins, with 102 large-scale (more than 60 MW) and 198 small-scale (less than 40 MW) sites [29]. The economically feasible estimated potential of Ethiopia is approximately 30,000 MW, but only 8.82 % of this potential is exploited[18][19][16]. This capacity is insufficient to meet Ethiopia's energy needs, given the country's hydropower potential. To meet energy demand in a variety of sectors, including agriculture, transportation, industries, and services, as well as to contribute to the nation's socioeconomic development. it is crucial to fully utilize the country's abundant hydropower resources. Based on this, the nation has started a plan to develop several hydroelectric projects, including the massive 6000MW GERD being built on the Blue Nile River and the 2,160 MW Gilgel Gibe IV hydropower project, to meet the nation's increasing energy demand. The plan also includes a plan to export electricity to nearby nations.

Ethiopia possesses significant hydropower resources that remain largely untapped. While the country has substantial hydroelectric potential, only a small fraction of it is currently being utilized. To address this, Ethiopia has initiated a strategic plan to develop several major hydropower projects, including the massive 6000MW GERD being built on the Blue Nile River and the 2,160 MW Gilgel Gibe IV hydropower project, and a plan to export electricity to nearby nations. These plans will help to meet the country's growing energy demand and contribute to its socio-economic development.

Hydropower remains a vital source of renewable energy globally. Accurately predicting hydropower production is crucial for optimizing grid operations, ensuring energy security, and facilitating water resource management. On the continental level (Africa), research on hydropower prediction has gained momentum, particularly in countries with substantial hydropower resources. (Here, you can insert specific examples of research conducted in African countries besides Ethiopia). In Ethiopia, hydropower represents the dominant source of energy generation. Several studies have explored various techniques for hydropower prediction in the country. However, there is a continuing need to investigate the efficacy of more advanced machine learning algorithms for enhancing prediction accuracy, particularly for medium-scale hydropower plants.

Several studies have investigated the use of various deep-learning models to predict hydropower production[2][21][26][29][32]. For example, by utilizing RNN, LSTM, and GRU models, it becomes possible to analyze large sets of hydro-logical data, including historical rainfall patterns, river flow rates, and climate variables. This analysis can help in predicting future water availability and optimizing the operation of hydroelectric power plants. These deep learning models can capture the temporal dependencies and complex relationships within the data, enabling more accurate forecasts and effective decision-making in hydropower management[2]. By incorporating socio-economic and geopolitical factors into the models, they can provide insights into the potential impacts of the dam on downstream countries and support the development of sustainable water-sharing agreements[2].

However, it is important to acknowledge the limitations of these deep learning models. One limitation is that they heavily rely on the availability and quality of data. In some cases, data may be limited or unreliable, which can affect the accuracy and generalizability of the models. Additionally, deep learning models can be computationally intensive and require substantial computing resources, which may pose challenges in resource-constrained environments.

Furthermore, the application of deep learning models such as RNN, LSTM, and GRU can significantly contribute to bridging the gap in Ethiopia's hydroelectric power utilization. These models enable the optimization of important information related to hydropower, facilitate the assessment of the energy industry's current status, and help resolve conflicts over water resources. However, it is essential to consider the limitations of these models, such as data availability and computational requirements, while leveraging their potential to achieve sustainable energy development in Ethiopia's Awash basin and beyond.

To address the existing gap in harnessing Ethiopia's hydroelectric potential, deep learning models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) were used to forecast future hydropower production.

1.2 General Objective

This thesis aims to develop and evaluate the effectiveness of recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and Gate Recurrent Units (GRUs) for predicting medium-scale hydropower production in Koka hydroelectric power plant, in Ethiopia.

1.3 Specific Objective

The specific objectives of this research are:

- To assess the suitability of RNNs, LSTMs, and GRU for medium-scale hydropower production prediction.

- To compare the performance of these machine learning models in predicting hydropower production.

- To identify the optimal model configuration for accurate prediction in the Ethiopian context.

- To analyze the factors influencing the prediction accuracy of the chosen model.

1.4 Statement of Problem

The demand for reliable models to predict hydropower production is increasing rapidly as energy consumption grows dramatically. The problem of quickly rising energy use in Ethiopia has led to the need to create trustworthy models to forecast hydropower production. These models will help electricity companies in Ethiopia manage and plan energy transmission, ensuring reliable and uninterrupted service to customers. Predicting hydropower is very important for electricity companies in Ethiopia to prepare short-term, medium-term, and long-term plans. The energy authority and government agencies also need to secure future energy supplies.
Forecasting hydropower production is useful not just for the power sector but can provide feedback that benefits all economic sectors in Ethiopia as they make plans for future development. Globally, predicting hydropower is difficult but important, as it is a key but uncertain factor for many countries trying to develop their electrical systems. In this research, the goal is to forecast short-term hydropower in Ethiopia using climate data and deep learning algorithms like long short-term memory (LSTM), gated recurrent unit (GRU), and recurrent neural network (RNN).

1.5 Research Question

This research will answer the following question, based on the predicting future hydropower production using climate data and discharge, the research questions accordingly:

- What deep learning model architecture provides the highest accuracy and lowest error rate in forecasting future hydropower production?
- Which optimization algorithm or hyperparameter tuning can be leveraged to achieve the best performance and accuracy in the hydropower forecasting models?
- Can the proposed model accurately predict the daily and weekly hydropower production in the proposed region under consideration?

- What are the key relationships and correlations between the climate variables (e.g., precipitation, temperature, wind speed, wind direction, and etc) and discharge measurements with the target variable of hydropower production?
- How does the predicted daily or weekly distribution of hydropower generation vary based on the climate and discharge patterns?

1.6 Significance of the study

Identifying the right deep-learning model for hydropower production prediction is crucial for optimizing energy management and maximizing efficiency. These models can analyze vast amounts of data, including historical generation data, climate, and discharge. By selecting the most effective model, hydropower plants can achieve more accurate predictions of future power output. This translates into several benefits: improved integration of hydropower into the electricity grid, enabling better grid balancing, and reducing the need for backup power sources. It also allows for more informed decisions about water releases from reservoirs, balancing power generation needs with flood control and environmental considerations. Additionally, accurate predictions empower hydropower producers to participate more effectively in electricity markets, by strategically buying and selling power based on anticipated generation. Generally, developing a reliable and accurate prediction model for medium-scale hydropower production is crucial for Ethiopia39;s power system stability and efficient water resource utilization. Thus, this study will contribute:

- Enhanced prediction accuracy for medium-scale hydropower plants.
- Improved grid operation efficiency through better forecasting of power generation.

2 Literature Review

The ever-growing demand for electricity pushes countries, especially developing nations to focus on finding dependable energy sources[9][23]. While increasing production capacity is essential, so too is refining power forecasting to ensure a stable supply [23][34]. Hydropower, a clean and renewable resource vital for combating climate change, offers a convincing solution. However, fluctuating demand patterns and the complex nature of hydropower production make accurate forecasting essential for optimizing output and bridging the gap between supply and demand [24]

Hydropower, a well-established and widely adopted renewable energy source derived from converting the energy of moving water into electricity, offers a compelling solution in the fight against climate change due to its minimal greenhouse gas emissions[25]. However, unlike controllable sources like coal or natural gas, hydropower production is heavily influenced by fluctuating river flows, reservoir volumes, temperature, and weather conditions [6]. This variability introduces significant challenges in accurately forecasting hydropower production, making precise predictions critical for optimizing output, meeting ever-growing energy demands, and ensuring grid stability[6]. Traditional forecasting methods, while previously employed, often struggle to capture the complex, non-linear relationships between these environmental factors and actual hydropower output [8]. This highlights the need for more sophisticated techniques, paving the way for this research to explore the potential of advanced machine learning models for accurate hydropower production forecasting.

Traditional forecasting methods, such as statistical models and rule-based approaches, have been historically employed for hydropower production prediction [16]. However, these methods often fall short of capturing the intricate and non-linear relationships between environmental variables like river flow, reservoir volume, weather patterns, and the actual output of hydropower plants. This inherent limitation restricts long-term forecasting accuracy, hindering effective energy planning and decision-making crucial for grid stability and sustainable resource management [7][31]. As a result, there is a growing need for more sophisticated forecasting techniques that can account for the complexities of hydropower production in an evolving environment.

In recent years, machine learning and deep learning techniques have emerged as powerful tools for tackling the inherent complexities of renewable energy forecasting [7]. These methods hold significant promise for capturing the non-linear relationships between various environmental factors and hydropower production, which is crucial for accurate predictions. Studies have successfully applied Long Short-Term Memory (LSTM) based deep learning models to forecast hydropower generation, demonstrating the effectiveness of incorporating historical data for improved long-term accuracy. Additionally, research has explored the use of Recurrent Neural Networks (RNNs) for short-term load forecasting, highlighting the potential of various network architectures in optimizing energy production forecasts [17]. This body of research paves the way for the development of more sophisticated machine-learning models specifically tailored to the unique challenges of hydropower production forecasting in developing nations.

Interestingly, there is a well-documented disparity in research focus between wind and hydropower forecasting, with wind power receiving considerably more attention despite hydropower's vital role in developing countries [26]. This is evident in studies comparing machine learning techniques for wind and hydropower forecasting, where wind power often takes center stage [13]. This imbalance highlights the need for further exploration and development of machine learning models specifically tailored to address the challenges and opportunities unique to hydropower forecasting, particularly in the context of developing nations. By focusing on hydro-specific models, we can address the complex interaction of factors influencing hydro production, such as river flow, reservoir capacity, and weather patterns, which are often not as prevalent a concern in wind forecasting [6]. This tailored approach has the potential to significantly improve forecasting accuracy in developing countries, where hydropower plays a critical role in ensuring stable and sustainable energy access.

Machine learning models have proven effective at predicting variables in engineering applications where the variable had been highly stochastic and complex to model using classical mathematical approaches. This study investigated the capability of various machine learning algorithms to predict power production from a reservoir in China. Specifically, the study evaluated the performance of three supervised and unsupervised machine learning models: Artificial Neural Network (ANN), AutoRegressive Integrated Moving Average (ARIMA), and Support Vector Machine (SVM). Among the evaluated models, the ANN had been the best performer, demonstrating the most accurate forecasting capabilities.

The projection of future hydropower generation is crucial for the sustainable development of countries that rely on hydropower as a major energy source. the impact of climate change has been investigated under two climate change scenarios and was tested for the future performance of the Samanalawewa hydropower plant in Sri Lanka using artificial neural networks (ANNs). The results showed that the forecasted hydropower generation would increase significantly, by 7.29 % and 10.22 %, under the two climate change scenarios tested[24]. Other researchers have applied various machine learning and deep learning models, including neural networks (ANNs), LSTM (long short-term memory), CNN (convolutional neural network), and hybrid CNN-LSTM designs, and integrated with meteorological data to produce real-time outcomes. This paper was employed to mitigate the error rate in the acquired outcomes, with comparisons conducted across light gradient boosting machines (LightGBM), gradient boosting regressor (GBR), and random forest regressor (RF) techniques, as well as the deep learning models. The favorable performance has been exhibited to enhance the accuracy of prediction LightGBM machine learning model, while the hybrid CNN-LSTM model had the greatest rate of inaccuracy[31].

In another paper, the researchers developed forecasting models that could accurately estimate electrical loads based on an electricity company's measurements of current electrical loads. The researchers had been employed three deep learning algorithms to forecast electrical loads: Long Short-Term Memory (LSTM), Gated

Recurrent Units (GRU), and Recurrent Neural Networks (RNN). After testing the models, the paper reported that the GRU model had been found to have the best performance, achieving an R-squared value of 90.228 , which indicated a high level of accuracy.[2].

This research investigates the potential of Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, for accurately predicting medium-term hydropower production in Ethiopia. RNNs have demonstrated success in time series forecasting tasks due to their ability to handle sequential data [13]. LSTM and GRU networks are particularly adept at capturing long-term dependencies within data, making them suitable for modeling the complex dynamics of hydropower production influenced by factors like reservoir volume, river flow, and weather [6][9]. By leveraging these machine learning techniques, we aim to identify a robust forecasting model that can outperform traditional methods which often struggle with non-linear relationships [10][28][29]. Furthermore, it should be noted here that we were not able to identify hydropower potential prediction in Ethiopia, particularly in the study region using state-of-the-art Artificial Intelligence methods (machine, and deep learning models). However, there is one study in the Omo-Gibe River basin using the HBV model [16]. These indicate the existing gap in using the models in predictions in any of the river basins in Ethiopia. Therefore, for the first time in the context of Ethiopia, hydropower prediction models such as RNNs, LSTM, and GRU are used. The findings further contribute to more efficient and sustainable energy management in Ethiopia, particularly as the country's installed hydropower capacity advances [26].

3 Data and Methodology

3.1 Study Area

The Koka Dam, nestled in Ethiopia's Awash basin, holds a distinguished position as one of the nation's oldest small-scale hydropower plants. Beyond generating electricity, this versatile infrastructure serves multiple purposes and is situated 75 kilometers southeast of Addis Ababa in the East Shewa Zone. The vast Awash basin, encompassing 110,000 km^2, is home to an estimated 10.5 million people. Strategically located at 8° 24'N latitude and 39° 5'E longitude, the Koka Dam is closely monitored by the Ethiopian Electric Power Corporation. The dam plays a critical role in the region's energy landscape, boasting an average daily power production of 279.589 MWh. However, hydropower production has been influenced by environmental factors. Consequently, understanding how climate variability affects these regions distinctively will help each of these countries for effective control of future hydropower production and decision-making. Figure 1 shows the geographical map of this positioning making it a valuable power generation asset for the region, contributing significantly to meeting its energy demands.

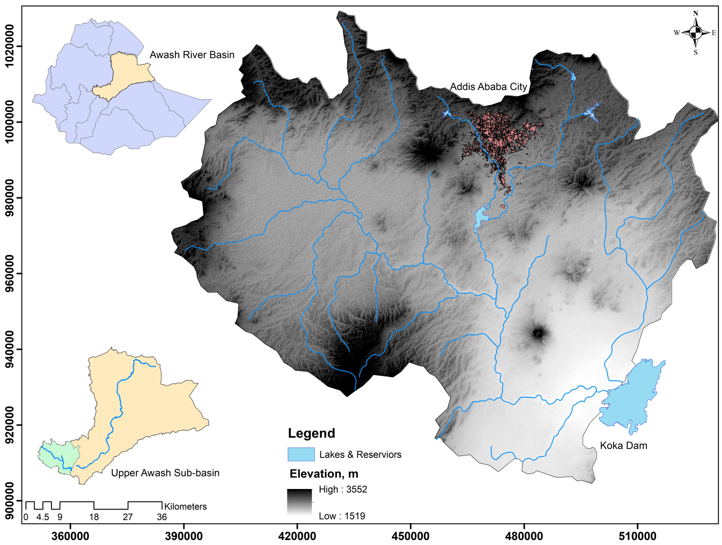

Figure 1: The Koka Dam location and surroundings.

3.2 Data Set

3.2.1 Climate Data

The climate data used in this analysis was collected from the NASA Power (Prediction of Worldwide Energy Resource) database. The NASA Power dataset contains high-quality, global-scale observational data spanning a 13-year period from 2010 to 2023. The daily records obtained from the NASA Power dataset including Precipitation (PRECTOTCORR), Minimum temperature (T2M-MIN), Maximum temperature (T2M-MAX), Relative humidity (RH2M), Maximum wind speed, Minimum wind speed, Wind direction, and All-sky surface shortwave radiation (ALLSKY-SFC-UV-INDEX)

3.2.2 Hydropower Production and Discharge

In addition to the climate variables, the dataset also includes hydropower production and discharge data for the Koka Dam in the study area. This research has considered both the daily and weekly records for the climate variables as well as the hydropower production and discharge data. The availability of daily and weekly reports of hydropower production from the Koka Dam was a key factor in the decision to utilize both daily and weekly temporal resolutions in the analysis. The comprehensive climate dataset, combined with the hydropower production and discharge data, will be instrumental in developing a robust model to understand the relationships between climatic factors and hydropower generation in the study area. The integration of these multi-year, multi-variable datasets will provide valuable insights that can inform energy planning and management strategies for the region.

3.2.3 Data Description

The dataset utilized in this study consists of 4,747 daily and 679 weekly records of hydropower generation data from the Ethiopia Electric Power Corporation, as well as climate data from 14 meteorological stations located near the Koka dam. For each of the 14 meteorological stations, the dataset includes 4,747 daily records covering the period from 2010 to 2023.

The dataset contains nine (9) independent variable attributes that represent various climate factors, and one (1) dependent target variable that represents the hydropower production. The independent variables include Precipitation (PREC-TOTCORR), Minimum temperature (T2M-MIN) maximum temperature (T2M-MAX), Relative humidity (RH2M), Maximum wind speed, Minimum wind speed, Wind direction, and Discharge. The target dependent variable is hydropower production, which reflects the increase or decrease in hydropower generation. Tables 1 2 in the paper present a sample of the daily raw dataset before any preprocessing or feature engineering was performed and Tables 4 and 5 sample the weekly raw data set. This raw data includes the values for each of the 9 independent variables and the corresponding hydropower production target variable. In addition, Table 3 represents a statistical summary of the daily raw data.

year	ALL-Sky-Sw-Dwn	PRCP	RHM	TMAX	TMIN
1/12/2010	4.8650	0.022778	73.933333	22.305556	10.862222
1/13/2010	5.1038	0.015000	66.981667	22.955556	10.729444
1/14/2010	5.9922	0.000000	61.245000	23.702222	8.731111
1/15/2010	6.1005	0.006111	54.234444	24.781667	10.216111
1/16/2010	6.5494	0.000000	51.028889	24.972778	10.426111
1/17/2010	6.1950	0.000000	56.652222	24.275556	8.856111
1/18/2010	6.0688	0.000000	60.914444	23.804444	8.436111
1/19/2010	5.8377	0.000000	58.875000	23.017222	8.738889
1/20/2010	5.8594	0.000000	59.946111	23.452778	9.022778
1/21/2010	5.9400	0.000000	60.783889	23.751667	8.253889

Table 1: Sample of the daily dataset before preprocessing

To ensure accuracy and reliability, this study utilized data from two sources: the Ethiopian Electric Power Corporation and NASA's website (https://power.larc.nasa.gov/). The data spanned 13 years, from September 2nd, 2010, to November 1st, 2023. This extensive timeframe captured variations in weather patterns, contributing to the robustness and accuracy of the trained models (RNN, LSTM, and GRU) for

year	WS-max	WS-min	WD	Discharge	hydropower production
1/12/2010	5.818889	3.469444	88.170000	2034171.36	217.326
1/13/2010	6.353333	2.711667	91.642778	2158902.72	230.652
1/14/2010	5.541111	1.573333	109.999444	1750432.32	187.012
1/15/2010	3.900000	0.637222	148.181111	2390394.24	255.384
1/16/2010	4.427222	1.140000	124.533333	2366919.36	252.876
1/17/2010	5.192222	1.467778	112.541667	1951232.40	208.465
1/18/2010	5.553889	2.149444	117.013889	1814548.32	193.862
1/19/2010	5.303889	1.947222	120.027778	1967200.56	210.171
1/20/2010	5.611111	1.378333	118.697222	1773972.72	189.527
1/21/2010	6.298333	1.533889	101.237778	1020698.64	109.049

Table 2: Sample of the daily dataset before preprocessing

	mean	std	min
ALL-Sky-Sw-Dwn	5.959e+00	1.031e+00	1.864
PRCP	2.960e+00	4.459e+00	0.000
RHM	6.483e+01	1.418e+01	13.273
TMAX	2.398e+01	2.487e+00	17.264
TMIN	1.145e+01	2.397e+00	2.831
WS-max	6.062e+00	1.374e+00	2.708
WS-min	2.296e+00	1.285086e+00	0.247
WD	1.240e+02	5.858148e+01	56.676
Discharge	2.616e+06	1.212e+06	0.000
hydropower production(MWH)	2.794963e+02	1.295353e+02	0.000000

Table 3: Table with daily dataset summary statistics

hydroelectric energy prediction. The chosen meteorological inputs for the models included precipitation (in millimeters), minimum and maximum temperatures (in degrees Celsius), relative humidity (as a percentage), maximum and minimum wind speeds (in meters per second), wind direction (in degrees), and all-sky surface shortwave downward irradiance (in KW-hr/m2). These factors – precipitation, minimum temperature, maximum temperature, relative humidity, maximum wind speed, minimum wind speed, wind direction, and all-sky surface shortwave downward irradiance – all significantly affect the behavior and amount of hydroelectric power generation.

Hydropower energy production data from September 2^{nd}, 2010, to November 1^{st}, 2023, is displayed in Figure 2 and Figure 3 on both a daily and weekly basis. Figure 2 (daily data) reveals significant fluctuations in Hydropower energy production, likely due to weather patterns and discharge variations. Similarly, the weekly data presented in Figure 3 also exhibits substantial week-to-week fluctuations. These fluctuations can be attributed to several factors, including snow flow patterns, snow melt, water storage and reservoir management, Hydrological conditions, environmental factors, and flow requirements.

year	ALL-Sky-Sw-Dwn	average-precip	Tmax-avr	Tmin-av	ws-max
2010-01-17	5.801019	0.007315	23.832222	9.970185	5.205463
2010-01-24	6.072222	0.000000	23.892778	8.404127	5.955794
2010-01-31	5.642778	0.197857	24.763968	10.878333	5.790635
2010-02-07	6.239841	0.009444	25.296984	8.589524	6.243651
2010-02-14	6.621587	0.102063	27.888492	11.526032	6.243651
2010-02-21	4.233889	5.268730	22.313175	13.355159	4.760397
2010-02-28	5.617063	2.475317	22.773968	12.457778	5.537857
2010-03-07	6.031905	0.796984	24.645794	12.302302	5.149444
2010-03-14	4.958333	5.605952	24.237857	13.184683	4.216984
2010-03-21	6.217937	2.132143	24.603254	11.299841	7.046270

Table 4: Sample of the weekly Data set before preprocessing

Table 6: These statistical properties indicate that the hydropower data may require careful analysis and the use of appropriate techniques to address the issues identified, such as non-normality, autocorrelation, and potential multicollinearity[3][33]. For time series data, the Durbin-Watson statistic is a useful test to detect the presence of autocorrelation in the residuals of a regression model. A Durbin-Watson statistic close to 0 indicates positive autocorrelation, while a value close to 2 indicates no autocorrelation.

year	ws-min	WD-aver	Discharge	hydropower production)
2010-01-17	1.833241	112.511389	2.108675e+06	225.285833
2010-01-24	1.675952	112.101667	1.851181e+06	197.775714
2010-01-31	2.464842	96.399286	3.761458e+06	401.865143
2010-02-07	3.835001	83.872063	5.538736e+06	591.745286
2010-02-14	1.914603	111.880635	6.174143e+06	659.630714
2010-02-21	1.605397	135.648651	5.814872e+06	621.247000
2010-02-28	1.852143	111.280714	4.008088e+06	428.214571
2010-03-07	1.723413	122.912063	4.355653e+06	465.347571
2010-03-14	0.965159	167.665159	4.355816e+06	465.365000
2010-03-21	0.075079	100.736429	4.285956e+06	457.901286

Table 5: Sample of the weekly Data set before preprocessing

Skew:	-0.244
Kurtosis:	3.151
Durbin-Watson:	0.023
Jarque-Bera (JB):	51.473
Prob(JB):	6.65e-12
Cond. No.	8.69e+07

Table 6: These statistical properties indicate that the hydropower data

In this case of the hydropower data, the Durbin-Watson statistic of 0.023 is very low, which suggests the presence of strong positive autocorrelation in the data.[3] This is an important observation, as the assumption of independent errors is often violated in time series data due to the inherent sequential and dependent nature of the observations. LSTMs and GRUs are two popular variants of RNNs that have been widely used for time series forecasting, sequence-to-sequence learning, and other time-dependent tasks. These models can effectively learn and capture the long-term dependencies in the data, making them a suitable choice for analyzing the hydropower data, which appears to exhibit strong autocorrelation.

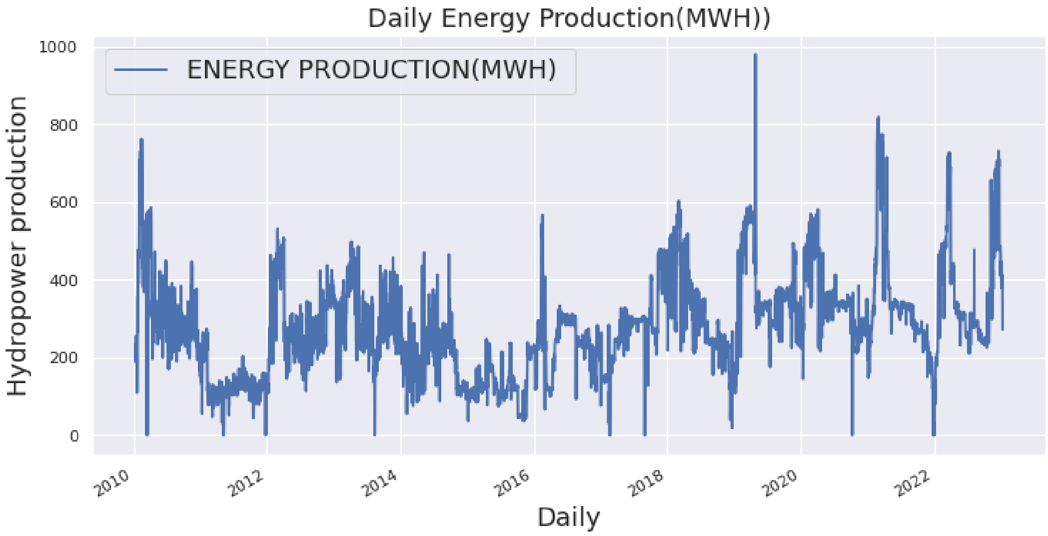

Figure 2: Daily time series Hydropower energy production average for specific domain over the period from 2010 to 2023.

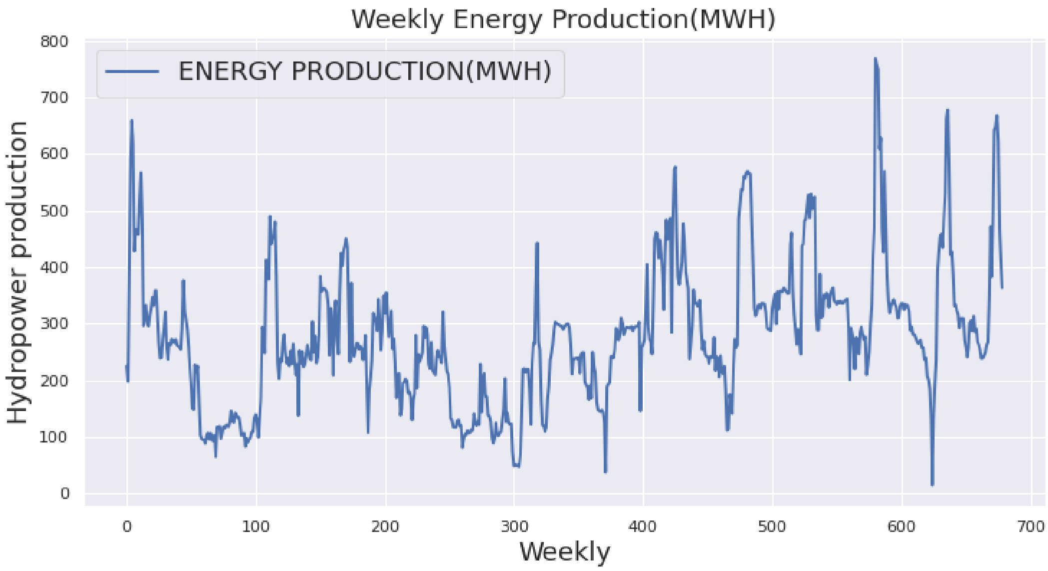

Figure 3: Weekly time series Hydropower energy production average for specific domain over the period from 2010 to 2023.

3.2.4 Data Pre-Processing

The dataset utilized in this study consists of a variety of continuous variables from heterogeneous sources, which can lead to issues with data quality and consistency. In the field of data mining and predictive modeling, ensuring high-quality data is a crucial step in achieving accurate and reliable predictions of future outcomes, such as hydropower production. To address the challenges posed by the inherent heterogeneity and potential inconsistencies in the dataset, the researchers applied several data preprocessing techniques:

- Missing Value Handling: The first step was to check the dataset for any missing values. Missing data can significantly impact the performance of machine learning models, as it can introduce bias and reduce the model's ability to capture the underlying patterns in the data. The researchers implemented the appropriate 3 missing value drops to ensure the dataset was complete and ready for further analysis.

- Outlier Detection and Handling: After careful analysis, the researchers determined that the identified outlier was deemed to be erroneous or irrelevant to the problem at hand. Therefore, the decision was made to remove the single outlier from the dataset. This step helps to improve the model's resilience to noisy or extreme observations, enhancing the overall predictive accuracy.

- Data Normalization: Given the heterogeneous nature of the variables, the researchers likely performed data normalization or scaling techniques to ensure that all features are on a similar numerical range as shown in Table 7 and Table 8. This step is important for many machine learning algorithms, as it can improve the convergence and stability of the model training process. The mathematical equation for Min-Max Normalization (Feature Scaling) is given as:-

$$x_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

year	ALLSKY-SFC-SW-DWN	PRCP	RHM	TMAX	TMIN
1/12/2010	0.471914	0.000428	0.796011	0.363222	0.597594
1/13/2010	0.509478	0.000282	0.704788	0.410055	0.587715
1/14/2010	0.649166	0.000000	0.629508	0.463854	0.439029
1/15/2010	0.666201	0.000115	0.537512	0.541630	0.549521
1/16/2010	0.736787	0.000000	0.495447	0.555400	0.565146
1/17/2010	0.681052	0.000000	0.569239	0.505164	0.448330
1/18/2010	0.661221	0.000000	0.625170	0.471219	0.417080
1/19/2010	0.624880	0.000000	0.598408	0.414498	0.439608
1/20/2010	0.628287	0.000000	0.612463	0.445881	0.460731
1/21/2010	0.640954	0.000000	0.623457	0.467417	0.403522

Table 7: Normalized Sample Data

3.2.5 Pearson's Statistical correlations

In this research paper involved the use of statistical correlation analysis to determine the degree of relationship between the relative movements of the climate variables and the hydropower production data. The formula for Pearson's correlation coefficient(r) is given as below.

$$r = \frac{\sum[(x_i - \overline{x})(y_i - \overline{y})]}{(n-1) \cdot \sigma_x \cdot \sigma_y} \tag{2}$$

Where:-

- x_i is the independent variable.
- y_i is the dependent variable.
- $\overline{x}$ is the mean of the independent variable
- $\overline{y}$ is the mean of the dependent variable
- σ_x is the standard deviation of the independent variable.
- σ_y is the standard deviation of the dependent variable.

year	WS-max	WS-min	WD	Discharge	Hydropower-produced
1/12/2010	0.360877	0.545574	0.130621	0.221753	0.221753
1/13/2010	0.422881	0.417270	0.145025	0.235350	0.235350
1/14/2010	0.328650	0.224532	0.221160	0.190821	0.190821
1/15/2010	0.138253	0.066033	0.379522	0.260586	0.260586
1/16/2010	0.199420	0.151162	0.281441	0.258027	0.258027
1/17/2010	0.288173	0.206660	0.231705	0.212711	0.212711
1/18/2010	0.330132	0.322077	0.250253	0.197811	0.197811
1/19/2010	0.301128	0.287837	0.262754	0.214452	0.214452
1/20/2010	0.336771	0.191515	0.257235	0.193388	0.193388
1/21/2010	0.416500	0.217853	0.184821	0.111270	0.111270

Table 8: Normalized Sample Data

- n is the number of observations.

This formula aims to help select the feature variable that has the strongest influence on hydropower production. Specifically, Pearson's correlation analysis was conducted to observe the strength of the association between the various climate variables and the target variable of hydropower production[14].

Pearson's correlation coefficient lies within the range of -1 to +1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and +1 indicates a perfect positive correlation[14]. The formula for Pearson's correlation coefficient is shown in Equation 2 of the paper. The goal of this feature engineering phase was to identify the climate variables that have the strongest influence on future hydropower production. By calculating the Pearson correlation coefficients between each climate variable and the hydropower target variable, the research team was able to determine which inputs would be most informative and impactful in the subsequent modeling process[14].

The use of Pearson's correlation analysis is a common and effective technique for feature selection.As it provides quantitative insight into the linear relationships between the predictor variables and the outcome of interest[21]. Identifying the climate variables with the highest correlation to hydropower output can help focus the model development on the most relevant and influential inputs, improving the overall predictive performance and interpretability of the final hydropower forecasting model.

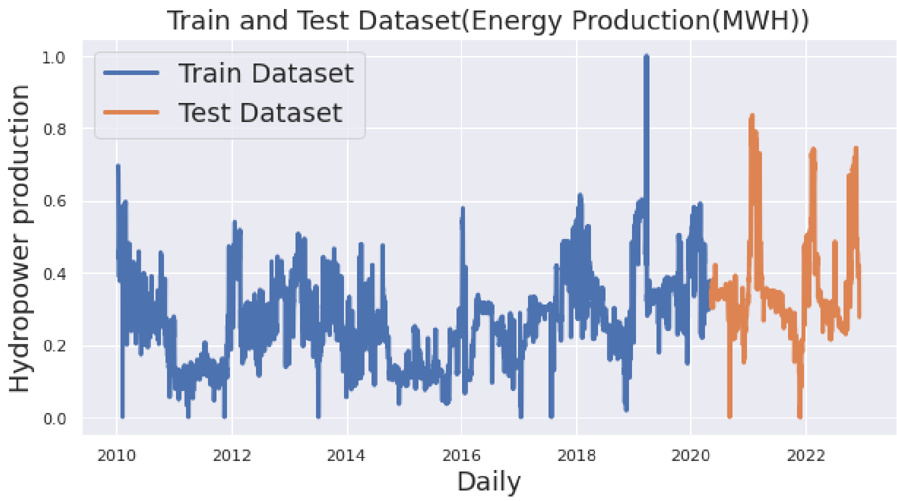

Figure 4: Hydropower Energy Production Train and Test Datasets with Different Splitting Strategy Daily split.

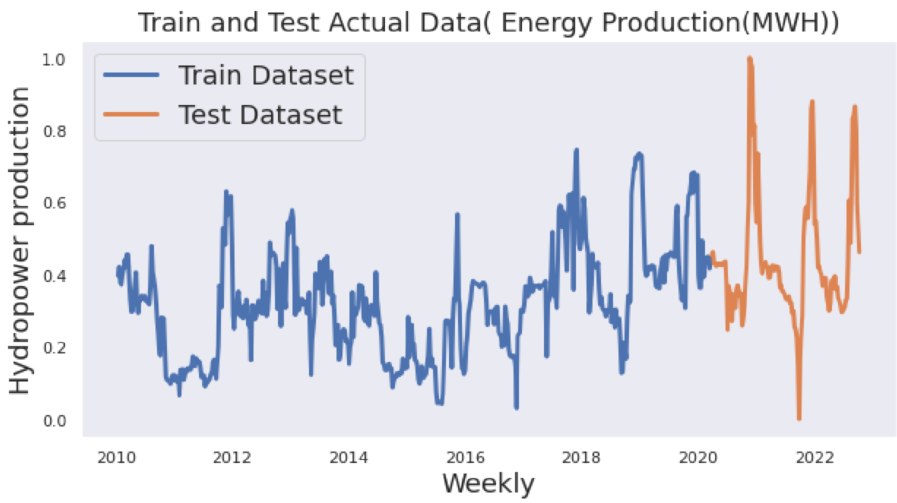

Figure 5: Hydropower Energy Production Train and Test Datasets with Different Splitting Strategy Weekly split.

Figure 6 presents a thermal map and a correlation graph to illustrate the relationships between input features and hydropower energy production. The thermal map uses a color scheme to visually represent the correlation coefficients calculated using the Pearson product-moment method for each feature. This method quantifies the correlation by dividing the co-variance between the target variables by their respective standard deviations[32]. The correlation graph ranges from +1 to -1. Where +1 signifies a perfect positive linear correlation, 0 indicates no linear correlation, and -1 signifies a perfect negative linear correlation[5]. The analysis reveals that average all-sky surface shortwave downward irradiance, maximum temperature, minimum temperature, maximum wind speed, minimum wind speed, and discharge all have a positive correlation with hydropower energy production. Conversely, precipitation, relative humidity, and wind direction exhibit a negative correlation with hydropower energy production.

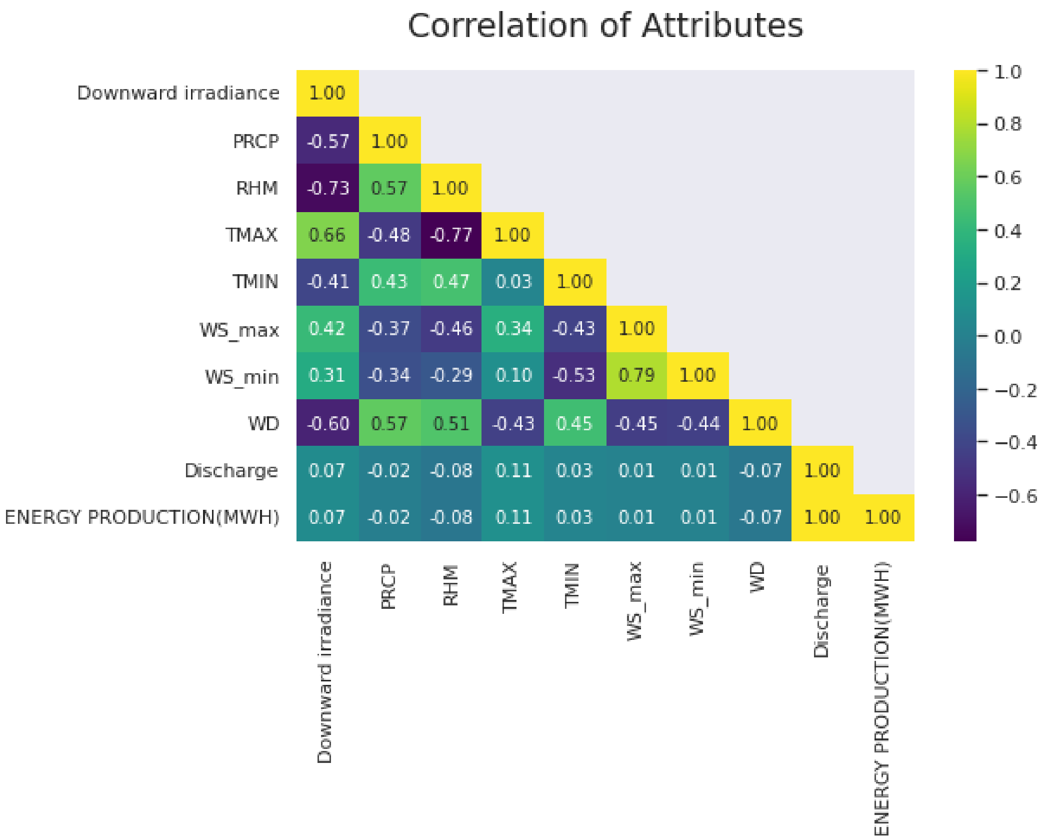

Figure 6: Correlation matrix showing the relationships between eight input variables and one output variable (production).

3.3 DEEP LEARNING ALGORITHMS

Deep learning, a subfield of machine learning, has revolutionized how machines process information. Inspired by the structure and function of the human brain, deep learning algorithms utilize artificial neural networks with multiple interconnected layers. These layers progressively extract and refine features from data, enabling the model to learn complex patterns and relationships. Unlike traditional machine learning methods, deep learning excels at handling vast amounts of unstructured data, such as images, text, and time series data. This capability makes deep learning a powerful tool for applications in diverse fields, from computer vision and natural language processing to healthcare diagnostics and, as explored in this thesis, hydropower production prediction.

3.3.1 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a specialized artificial neural network designed to process sequential data[4] as shown in Figure 7. Like a traditional feed-forward neural network, an RNN consists of an input layer, a hidden layer, and an output layer as shown in Figure 8. [2] [12] The key difference lies in the structure

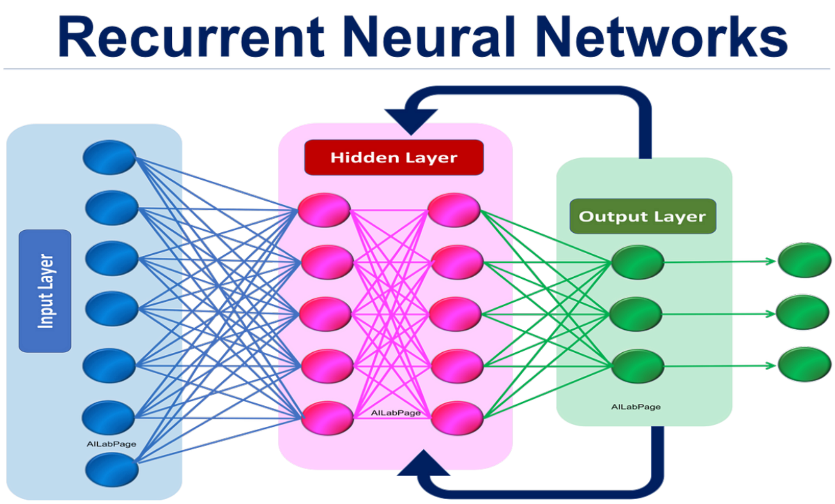

Figure 7: The RNN architecture consists of essential components, including input gates and the memory cell, which work together to process and store information to handle sequential data effectively.

of the hidden layer. In a feedforward neural network, the nodes in the hidden layer are only connected to nodes in the adjacent layers, with no connections between nodes in the same hidden layer. In contrast, an RNN has recurrent connections within the hidden layer, where the nodes in the hidden layer are connected. This recurrent structure allows the RNN to maintain an internal state or memory, which is crucial for processing sequential data. The mapping of the hidden state gt at the current time step t can be represented as:

$$g_t = f(UX_t + W.g_{t-1}) \tag{3}$$

- x_t is the input at the current time step
- g_{t-1} is the hidden state from the previous time step
- U and W are the learnable weight matrices that connect the input and previous hidden state to the current hidden state
- f is a nonlinear activation function such as tanh or ReLU, that transforms the weighted sum of the inputs and previous state into the current hidden state

This recurrent computation enables the RNN to incorporate information from past inputs and states, allowing it to capture and model patterns in sequential data. Maintaining a state memory and propagating information through time is a key strength of RNNs, making them particularly useful for tasks like language modeling, speech recognition, and time series prediction[2] [12]

3.3.2 Long Short-Term Memory (LSTM)

LSTM networks, a type of Recurrent Neural Network (RNN), are specifically designed to address the vanishing gradient problem and capture long-range relationships within sequential data[4]. The Long Short-Term Memory (LSTM) network has a specialized structural component called a gate. Their structure is more complex compared to GRU networks.[4] LSTMs utilize three gates (input, forget, and output) to control the flow of information in and out of the memory cell (as illustrated in Figure 8).

This gate takes an input vector and produces an output value that ranges from

0 to 1. The purpose of this gate is to control the flow of information through the LSTM network. When the gate output value is 0, it means no information is

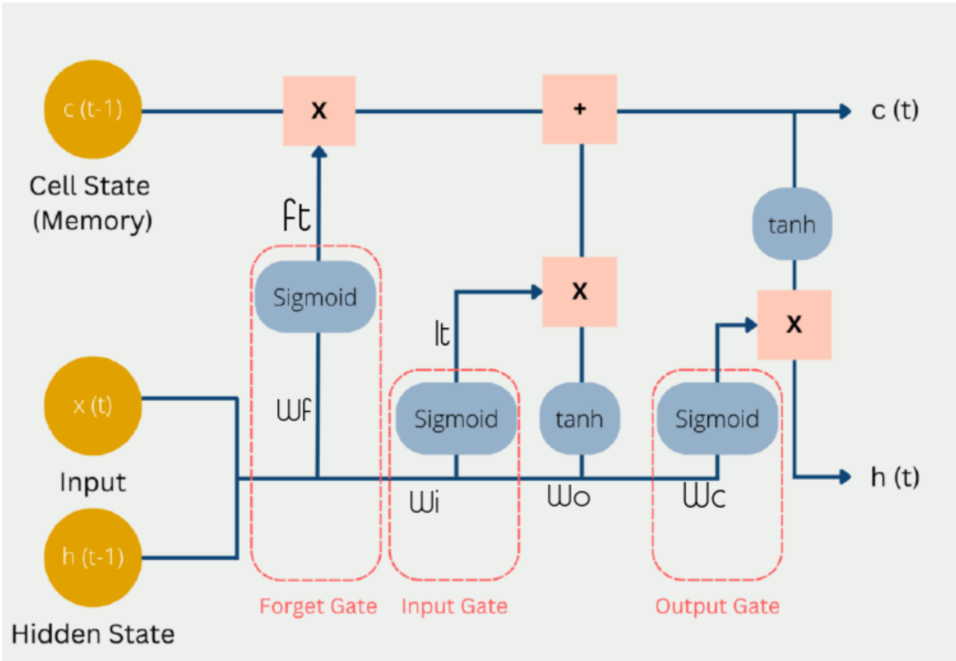

Figure 8: The LSTM architecture comprises the essential components of an LSTM cell, such as input, forget, and output gates, along with the memory cell, which work together to process and store information over time.

allowed to pass through. Conversely, when the output value is 1, all information can pass through. Given the current input vector x = x_1, x_2, ..., x_{t-1}, x_t and the previous output vector s = s_1, s_2, ..., s_{t-1}, s_t, the calculation formula for the LSTM gate can be expressed as follows:-

$$g(x) = \sigma(wx - b) \tag{4}$$

where:-

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

- w is the weight matrix;
- b is the bias vector

In a Long Short-Term Memory (LSTM) network, the cell state, denoted as ct, serves as the core component responsible for maintaining and updating the current state of the LSTM. It acts as a memory bank that stores the relevant information from the past. The updated cell state can be computed using the following formula:

$$c_t = f_t \times c_{i-t} + i_t \times \tanh(w_c[s_{t-1}, x_t] + b_c \tag{6}$$

Where:-

- w_c is the weight matrix of the cell state
- b_o is the bias vector of the cell state
- i_t means the input gate, which determines how much of the input at the current time is saved in the cell state;
- f_t represents the forget gate used to help the network Foregate the past input information and reset the memory cells.

The calculation of the input gate and forget gate can be respectively expressed as:-

$$i_t = \sigma(w_i.[s_{t-1}, x_t] + b_i) \tag{7}$$

$$f_t = \sigma(w_f.[s_{t-1}, x_t] + b_f) \tag{8}$$

where:-

- w_i is the weight metrics of the input gate
- w_f is the weight metrics of the forget gate
- b_i is the bias vectors of the input gate
- b_f is the bias vectors of the forget gate

In a Long Short-Term Memory (LSTM) network, the output gate plays a crucial role in controlling the flow of information from the current cell state to the current output. The output ht at the current time step can be expressed as:

$$O_t = \sigma(W_o.[s_{t-1}, x_t] + b_o) \tag{9}$$

where:-

- W_o is the weight matrix in the output gate and
- b_o is the bias vector in the output gate

The final output of the LSTM is computed as:

$$s_t = O_t \times tanh(t) \tag{10}$$

While training LSTMs is computationally expensive compared to GRUs, they can perform better in tasks requiring capturing long-term dependencies. Stacking multiple LSTM layers together can further improve performance on complex tasks. LSTMs have been successfully applied in various domains such as healthcare, finance, and robotics. Their applications extend to language identification, rainfall-runoff modeling, and hydrological concept formation [1].

3.3.3 Gated Recurrent Unit (GRU)

GRU is a newer and more efficient version of the Long Short-Term Memory (LSTM) architecture, which is a type of recurrent neural network (RNN)[12]. GRUs are designed to address the long-term dependency learning problems that can arise in traditional RNNs as shown in Figure 9. Compared to LSTMs, GRUs have a simpler architectural structure, with only two gating mechanisms:

- **Update Gate (z_t):** This gate determines what information from the previous hidden state (h_{t-1}) should be remembered and carried forward to the current hidden state (h_t). It combines the functionality of the input and forgets gates in LSTMs.
- **Reset Gate (r_t):** This gate controls how much of the previous hidden state (h_{t-1}) should be remembered when computing the current hidden state (h_t). It decides which information from the past should be discarded or retained, allowing the GRU to capture short-term dependencies in the sequence data.

Unlike LSTMs, GRUs do not have a separate cell state. The update gate is responsible for managing the flow of information between the previous hidden state and the current hidden state. The mathematical representation of the GRU operations can be defined as[34][**?**]:

$$r_t = (Wr * x_t + Ur * h_{t-1}) \tag{11}$$

$$z_t = (Wz * x_t + Uz * h_{t-1}) \tag{12}$$

$$h_t = tanh(W * x_t + U * (r_t * h_{t-1})) \tag{13}$$

Where:

- σ is the sigmoid activation function

The simpler structure of GRUs, with fewer parameters compared to LSTMs, often leads to faster training times. Additionally, the gating mechanisms in GRUs help mitigate the vanishing gradient problem that can occur in traditional RNNs, making GRUs more effective in learning long-term dependencies in sequence data.

3.4 Evaluation Metrics

This paper utilized four key evaluation metrics to assess the performance of the GRU, LSTM, and RNN models for weekly hydropower prediction:

- Root Mean Squared Error (RMSE): RMSE is calculated by taking the square root of the average squared differences between the predicted and actual values. It provides a measure of the average magnitude of the errors, giving a sense of how close the predicted values are to the true observed values. RMSE is useful for comparing model performance across different datasets or modeling problems, as it is in the same units as the original target variable.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |y_i - y_{pred}|^2} \tag{14}$$

- Mean Absolute Error (MAE): MAE measures the average of the absolute differences between the predicted and actual values. Unlike RMSE, MAE does not square the errors, so it is less sensitive to outliers. MAE provides a robust and interpretable measure of average model error, representing the typical magnitude of errors in the same units as the target variable.

$$MAE = \sum_{i=1}^{N} |y_i - y_{pred}| \tag{15}$$

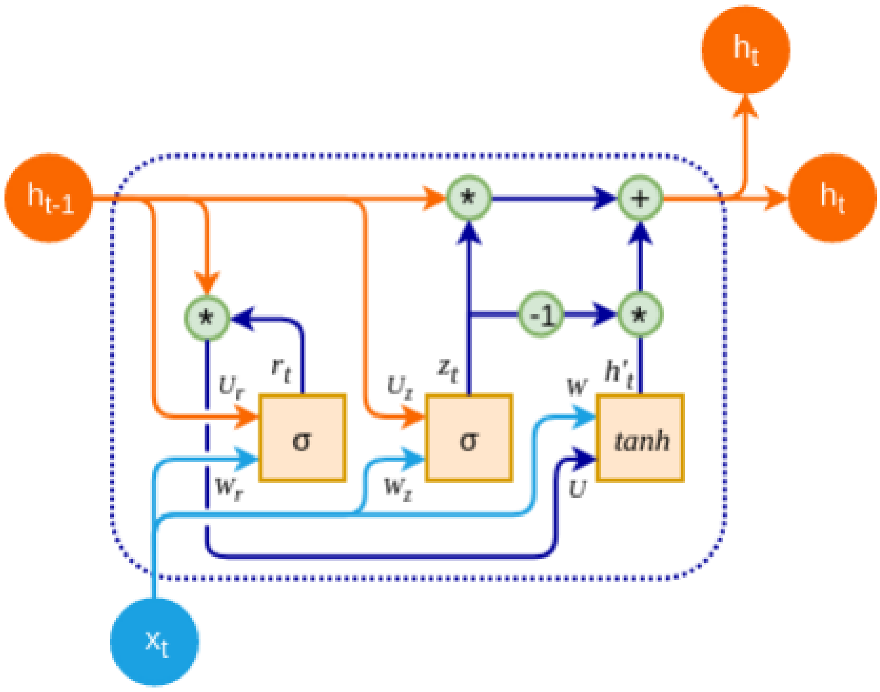

Figure 9: The internal mechanisms of the GRU architecture demonstrate how it effectively captures and retains important information when dealing with sequential data.

- Mean Squared Error (MSE): MSE calculates the average of the squared differences between the predicted and actual values. By squaring the errors, MSE places a higher penalty on larger errors, making it more sensitive to outliers compared to MAE. MSE is a useful metric for optimizing model parameters during the training process, as it can be directly minimized.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y_{pred}|^2 \tag{16}$$

- R-squared : R-squared is a statistical measure that represents the proportion of the variance in the dependent variable (in this case, the weekly hydropower generation) that is explained by the independent variables (the GRU, LSTM, or RNN model predictions). R-squared values range from 0 to 1, with 1 indicating that the model explains all of the variability in the target variable. R-squared is a valuable metric for assessing the overall goodness of fit of a model and its ability to capture the underlying patterns in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}|y_i - y_{pred}|^2}{\sum_{i=1}^{N}|y_i - y_{pred}|^2} \tag{17}$$

These four metrics provide a comprehensive quantitative assessment of the model's accuracy, precision, and predictive power. They allow for a thorough comparison of the GRU, LSTM, and RNN models, enabling the identification of the most suitable model for weekly hydropower forecasting applications based on the specific performance requirements and characteristics of the problem domain[12] [1].

3.5 System implementation

The process of this work mentioned in Figure 10 illustrates the series of steps involved in hydro-power prediction using RNN, LSTM, and GRU models. The evaluation of the deep learning-based predictive model was implemented using Anaconda 3, which supports the Python 3.6 programming language. The model was developed and tested using the PyTorch API, a popular open-source machine-learning library for Python. Some package libraries that support deep learning used for this study include Numpy, Pandas, Seaborn, Matplotlib, Datetime, Torch, and Adfuller test. The parameters that are used in this study to evaluate the models are given in Table 10. The workflow for hydropower production prediction

Parameters	Number of Used
Hidden size	90
Number of layers	3
Learning rate	0.001
Number of epochs	600
optimizer	Adam

Table 9: The key parameters used for the RNN, LSTM, and GRU models:

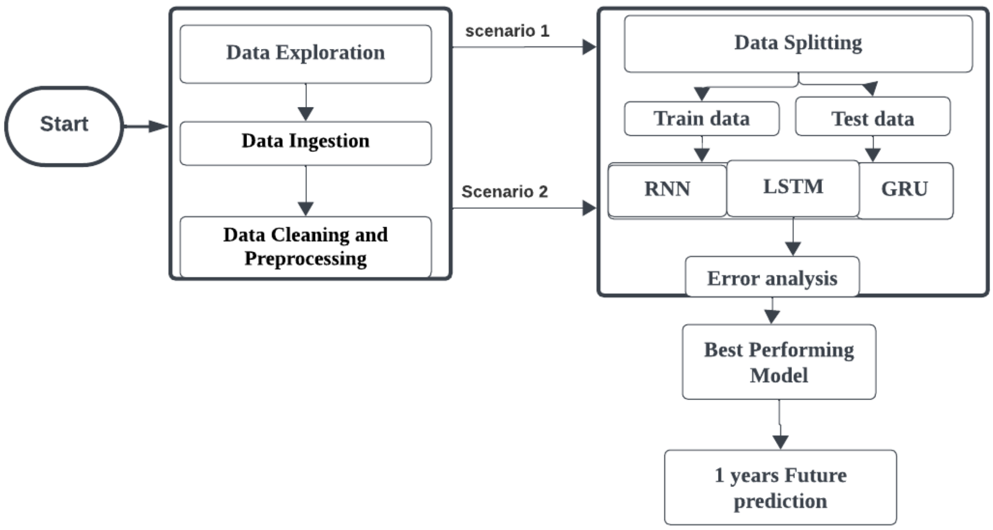

Figure 10: Workflow for Hydropower Prediction with RNN, LSTM, and GRU Models.

using RNN, LSTM, and GRU models will involve several key steps. First, historical data encompassing climate information, reservoir levels, turbine operations, and past power plant output will be collected and preprocessed. This preprocessing may involve data cleaning, normalization, and potentially feature engineering to ensure compatibility with the deep learning models. Next, the RNN, LSTM, and GRU architectures will be implemented, with each model being trained on the prepared historical data. The training process will involve iteratively adjusting the models' internal parameters to optimize their ability to learn the underlying relationships between input features and hydropower production. Once trained, the performance of each model will be evaluated on a separate validation dataset. This evaluation will assess the models' accuracy in predicting hydropower production. Finally, the best-performing model will be selected and employed for real-time or short-term forecasting of hydropower production at the Koka Dam Hydropower Plant.

4 Results and Interpretations

4.1 Train Test split on Daily and Weekly bases

In this study, RNN, LSTM, and GRU algorithms, which are well-known for their effectiveness in analyzing sequential data, were used to develop models for predicting electric power generation. The data set was divided into two subsets: a training set comprising 80 % of the data and a test set with the remaining 20 % as shown in Figures 9 and Figure 10).

Figure 9 shows the daily hydropower energy production data divided into training and testing sets. The training data (blue line) is used to build a model for predicting future production, while the test data (red line) is used to evaluate the model's performance. Similarly, figure 6 depicts the same data split into training and testing sets weekly. Similar to the daily split, the training data (blue line) is used for model development, and the testing data (red line) is used for model evaluation.

4.2 Training and Validation Loss for Daily and Weekly Dataset and each model

The training aimed to minimize the difference between predicted and actual hydropower production values. Three loss functions were employed to achieve this: R-squared (R^2), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The models were rigorously trained over 600 epochs, with each epoch accurately iterating through the entire dataset. A batch of data points was processed together during each iteration, and a fixed learning rate of 0.001 ensured controlled learning. To optimize the models and accelerate convergence, the Adam optimizer, known for its adaptive momentum estimation capabilities was utilized by manual hyperparameterization.

The trained models were assessed using the reserved test dataset for daily predictions following the training process. To gain valuable insights into the models' performance and ability to learn daily, the study visualized the training and testing loss across epochs for the LSTM, GRU, and RNN models (illustrated in Figure 11). These graphs specifically show how the loss values change over each epoch, providing crucial information about how well the models adapt and improve when predicting daily hydropower production. By analyzing these visualizations, we can effectively evaluate the accuracy and reliability of these models for short-term electric power forecasting.

Similar to the daily evaluation, the trained models were assessed using the reserved test dataset for weekly predictions. To understand the models' performance and ability to learn every week, the study visualized the training and testing loss across epochs for the LSTM, GRU, and RNN models (Figure 12). These graphs provide crucial information about how well the models adapt and improve when predicting weekly hydropower production. By analyzing these visualizations on a weekly time scale, we can effectively evaluate the accuracy and reliability of these models for medium-term electric power forecasting.

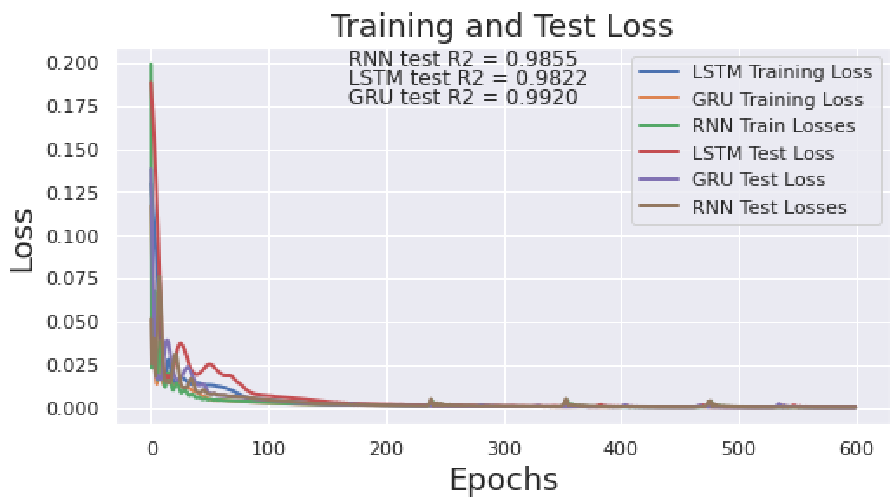

Figure 11: Training and test loss of the electric power prediction models on different time scales Daily.

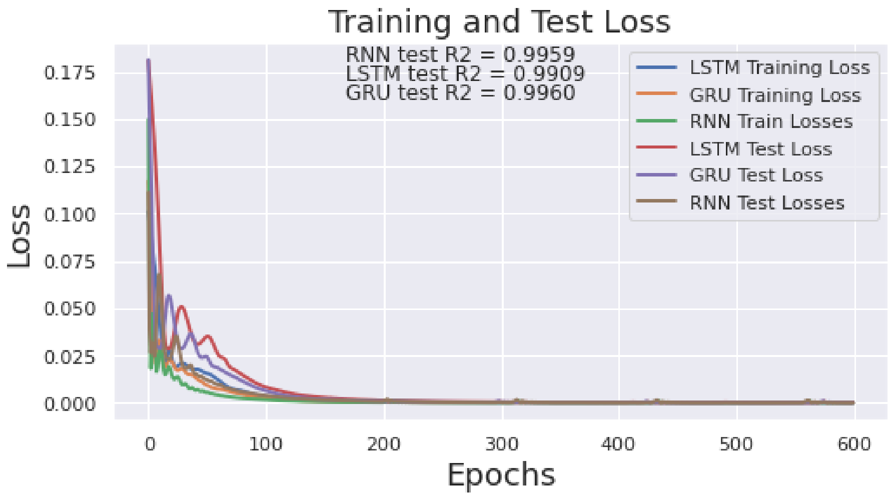

Figure 12: Training and test loss of the hydropower production prediction models on different time scales Weekly.

4.3 Daily Hydropower Production Prediction Models Result

Following the train-test split the daily dataset was trained on the training dataset (80 %) and each model was tested on the remaining 20 % dataset and tested the predicted data set. The findings are displayed in Figure 13, Figure 14, and Figure 15 for each of the three models.

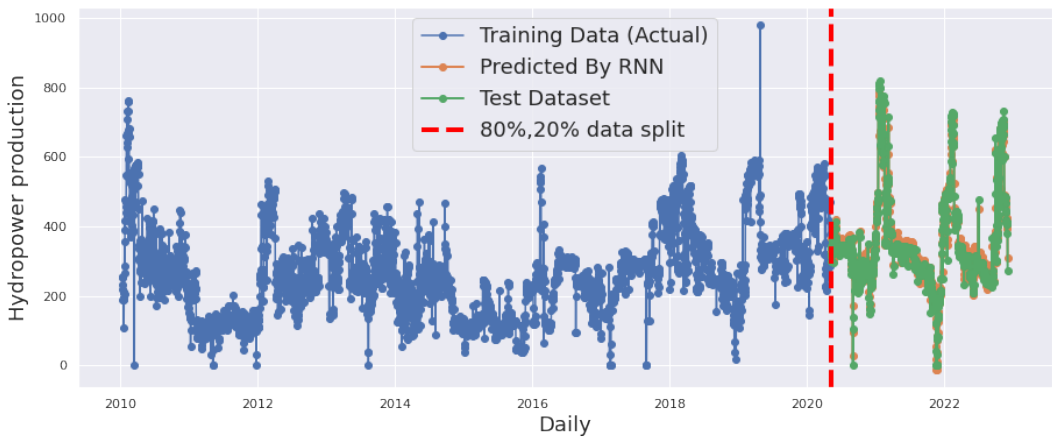

Figure 13: Comparison of actual and predicted hydropower production daily using RNN model.

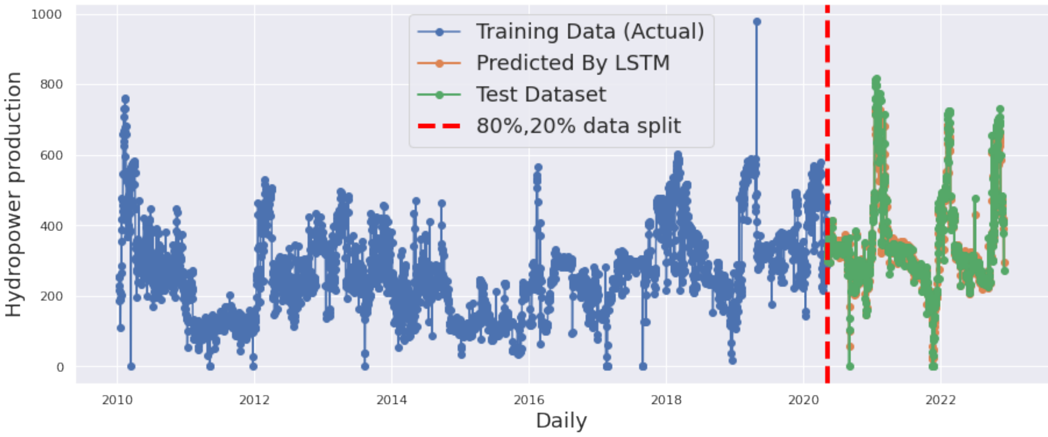

Figure 14: Comparison of actual and predicted hydropower production daily using LSTM model.

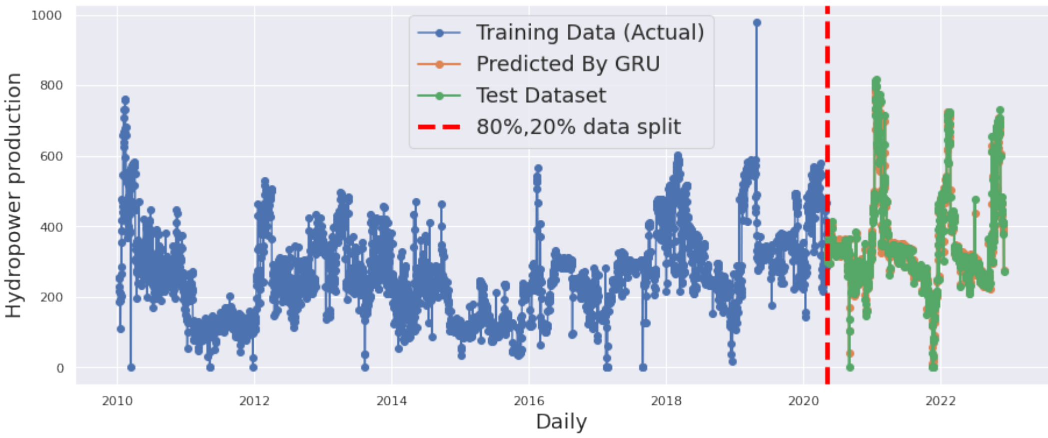

Figure 15: Comparison of actual and predicted hydropower production daily using GRU model.

4.4 Weekly Hydropower Production Prediction Models Result

Following the train-test split the weekly dataset was trained on the training dataset 80 % and each model was tested on the remaining 20 % dataset and tested predicted. The findings are displayed in Figure 14 Figure 15, and Figure 16 for each model. Plotting the test data and its prediction separately, the daily actual

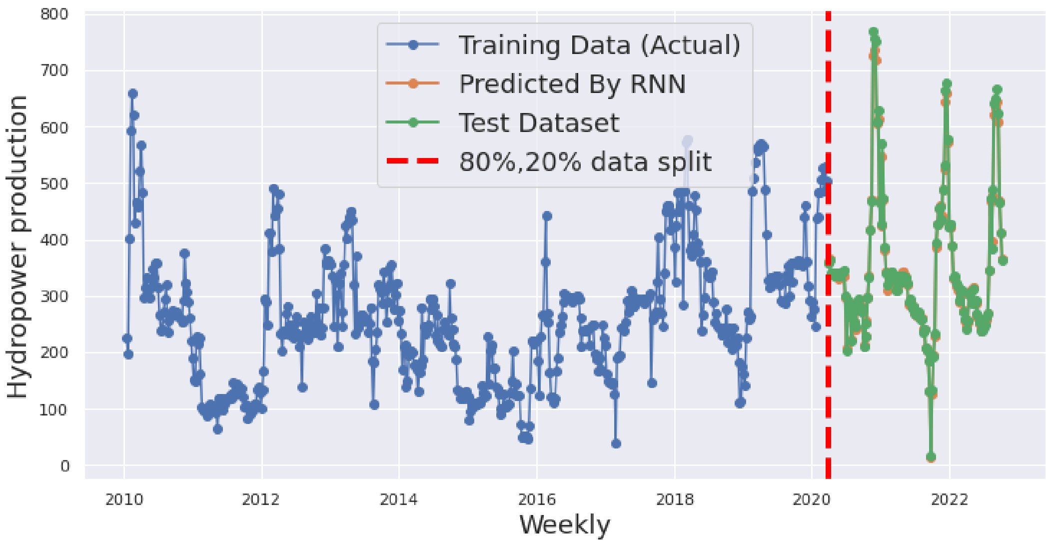

Figure 16: Comparison of actual and predicted hydropower production weekly using RNN model.

(daily test dataset) and prediction made using RNN, LSTM, and GRU are shown in Figure 19. Similarly, the weekly actual (weekly test dataset) and prediction made using RNN, LSTM, and GRU are shown in Figure 20.

4.5 Model Comparison

In addition to a graphical display of the results presented in section 4, the study thoroughly evaluated the performance of the RNN, LSTM, and GRU models for both daily and weekly hydropower production predictions. The researchers utilized common evaluation metrics to assess the models' accuracy and reliability. These performance metrics include R-square, MSE, MAE, and RMSE. Table 11 in the study displays the performance evaluation metrics for the three models in the daily hydropower prediction scenario. The results show that the proposed GRU

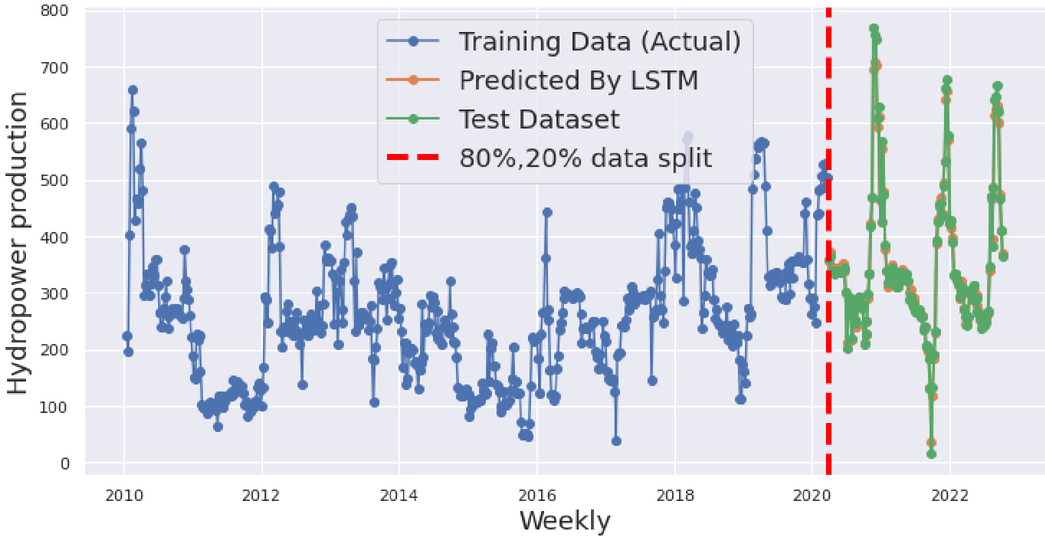

Figure 17: Comparison of actual and predicted hydropower production weekly using LSTM model.

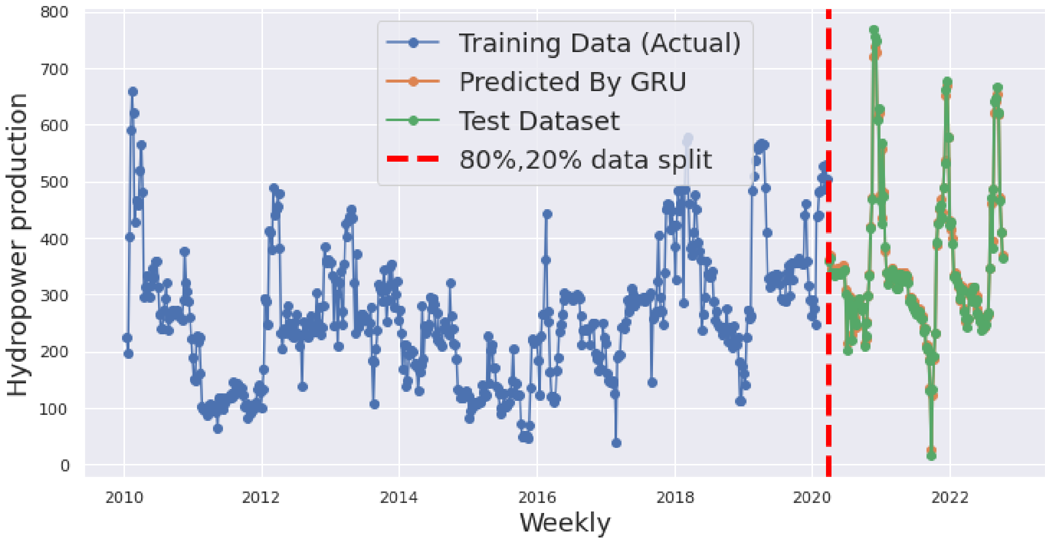

Figure 18: Comparison of actual and predicted hydropower production weekly using GRU model.

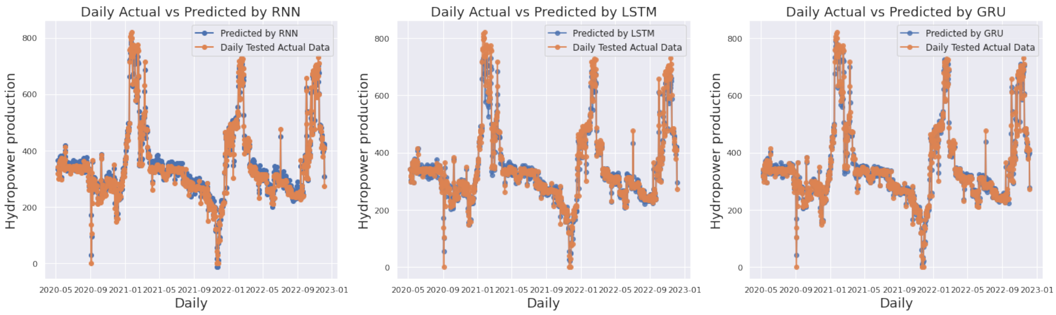

Figure 19: Predicted test and actual test dataset for the RNN, LSTM, and GRU models of daily scenarios.

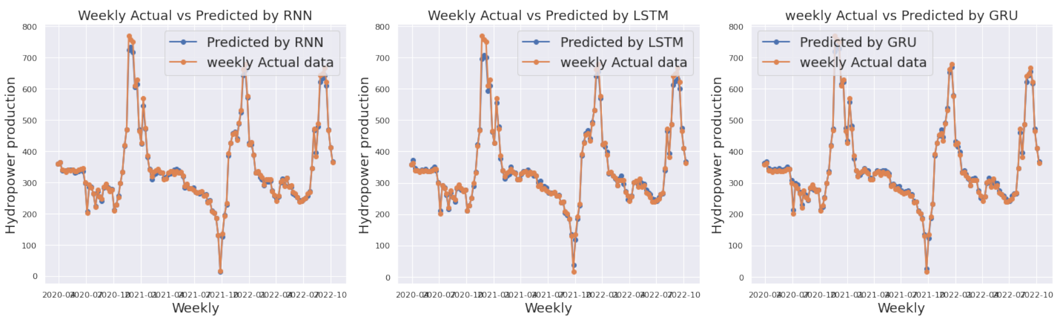

Figure 20: Predicted test and actual test dataset for the RNN, LSTM, and GRU models of daily scenarios.

model outperformed the LSTM and RNN models across all four evaluation metrics. Specifically, the four evaluation metrics (R-squared, MSE, MAE, and RMSE) of the GRU model were smaller than those of the other two models. The GRU model's performance was able to improve by at least 0.98% compared to the LSTM model and 0.65% compared to the RNN model in terms of these four metrics. The study's

Model/Performance	R-squared	MSE	MAE	RMSE
GRU	0.9920	143.55	8.612	11.981
LSTM	0.9822	317.88	11.568	17.829
RNN	0.9855	259.64	12.508	16.113

Table 10: Performance of three models in a daily hydropower prediction.

comparison of the daily hydropower prediction results provides valuable insights into the comparative performance of the RNN, LSTM, and GRU models. The researchers compared the predicted daily hydropower values generated by each model against the actual values from the dataset and visualized the results in Figure 6. The visual representation in these figures demonstrates that all three recurrent neural network architectures - RNN, LSTM, and GRU - can closely match the actual daily hydropower production patterns. This indicates that these models are capable of effectively capturing the complex dynamics and temporal dependencies inherent in hydropower energy generation. However, a more detailed comparative analysis of the models' performance reveals that the GRU (Gated Recurrent Unit) model outperforms both the RNN and LSTM models in the task of daily hydropower prediction. This superiority is evident from the GRU model's significantly lower mean squared error (MSE) and substantially higher coefficient of determination (R-squared) values. Specifically, the GRU model achieves an impressive R-squared score of 0.9920, which is considerably higher than 0.9822 for the LSTM model and 0.9855 for the RNN model. This indicates that the GRU model's predicted values exhibit a much stronger alignment and better fit with the actual hydropower production data, compared to the other two recurrent neural network architectures. Several key factors contribute to the GRU model's superior performance in this application:

- Improved ability to capture long-term dependencies: The GRU architecture is particularly well-suited for learning and leveraging long-term temporal dependencies in the hydropower energy production data. This allows the model to better capture the complex, non-linear patterns and seasonal fluctuations inherent in hydropower generation.

- Computational efficiency: The GRU model is generally more computationally efficient compared to the LSTM model, as it has a simpler structure with fewer parameters. This efficiency enables the GRU to train more quickly and operate with lower resource requirements, without sacrificing predictive accuracy.

- Effective feature extraction and learning: The GRU model's unique gating mechanisms and recurrent structure enable it to effectively extract relevant features and learn the underlying relationships within the hydropower dataset. This results in more accurate predictions compared to the RNN and LSTM models.

By leveraging the superior performance of the GRU model in predicting daily hydropower production, energy companies can benefit from these precise forecasts. Accurate short-term hydropower predictions can help optimize energy generation strategies, improve resource allocation, and enhance overall operational efficiency in the hydropower sector. This can lead to significant cost savings, increased energy reliability, and better integration of hydropower into the broader energy grid.

In general, the study's comprehensive evaluation of the RNN, LSTM, and GRU models, supported by detailed performance metrics and visual analysis, provides a strong evidence-based foundation for the selection and implementation of the GRU model for daily hydropower forecasting applications.

The performance evaluation metrics for the three models (GRU, LSTM, and RNN) in the weekly hydropower prediction scenario are presented in Table 12. The results indicate that all three neural network architectures were able to perform well in predicting the test data, as evidenced by their high R-squared (R-squared) scores and relatively low prediction error metrics. Despite the overall strong performance across the board, a closer examination of the evaluation metrics reveals some noteworthy differences between the models.

The GRU (Gated Recurrent Unit) model demonstrated the highest R-squared

Model/Performance	$R_{squared}$	MSE	MAE	RMSE
GRU	0.9960	66.219	5.992	8.137
LSTM	0.9909	152.09	7.22	12.33
RNN	0.9959	67.887	5.22	8.239

Table 11: Performance of three models in a weekly hydropower prediction.

value of 0.9960, showcasing its ability to produce predictions that align most closely with the actual weekly hydropower production values. Furthermore, the GRU model also exhibited the lowest Mean Squared Error (MSE) of 66.219 and the lowest Mean Absolute Error (MAE) of 5.992 indicating that it generated the most accurate predictions with the smallest deviations from the ground truth. In comparison, the LSTM (Long Short-Term Memory) model achieved an R-squared value of 0.9906, which is slightly lower than the GRU's performance. The LSTM's MSE and MAE metrics of 152.09 and, 7.22, respectively, were also higher than those of the GRU model.

The RNN (Recurrent Neural Network) model, while still performing well, had the lowest R-squared value of 0.9959 among the three architectures. Its MSE of 67.882 and RMSE of 8.239 were slightly higher than the GRU's, but lower than the LSTM's. Based on these evaluation metrics, the GRU model can be considered the top performer in the weekly hydropower prediction scenario, outperforming the LSTM and RNN models by at least 0.5 % and 0.01 %, respectively, in terms of the R-squared value.

It is important to note that while the performance differences between the models may seem relatively small, the choice of the best model for a particular application can also depend on other factors beyond these four evaluation metrics, such as computational efficiency, training time, and interpretability of the model.

Nevertheless, the superior performance of the GRU model, as evidenced by its higher R-squared value and lower prediction errors, suggests that it is a slightly suitable choice for weekly hydropower forecasting applications. By leveraging the GRU model's ability to accurately predict weekly hydropower production, energy companies can make more informed decisions, optimize their generation strategies, and enhance the overall efficiency and reliability of their hydropower operations.

4.6 Future Hydropower Production Estimate

Accurate one-year predictions of daily and weekly scenarios using the outperforming GRU model are shown in Figures 21 and 23. The separate plot for one-year daily hydropower prediction is shown in Figure 22. Similarly, one-year weekly estimates are shown in Figure 24.

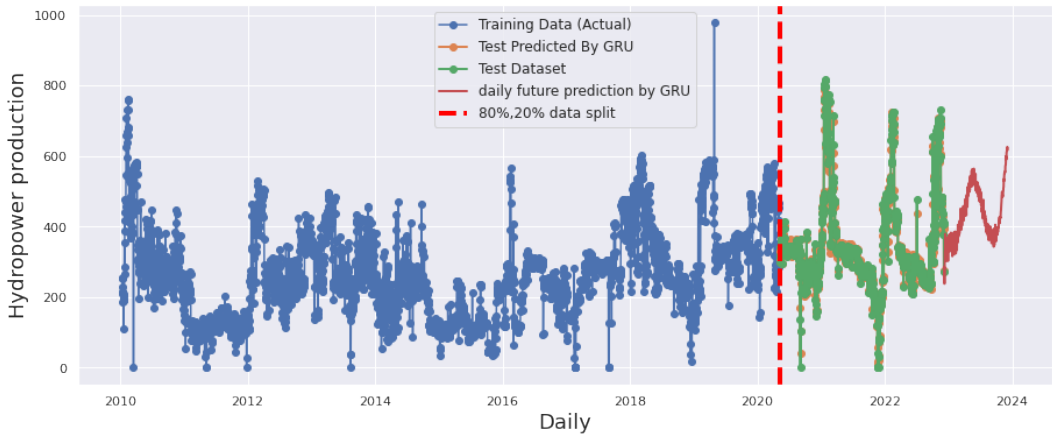

Figure 21: Daily Future Prediction.

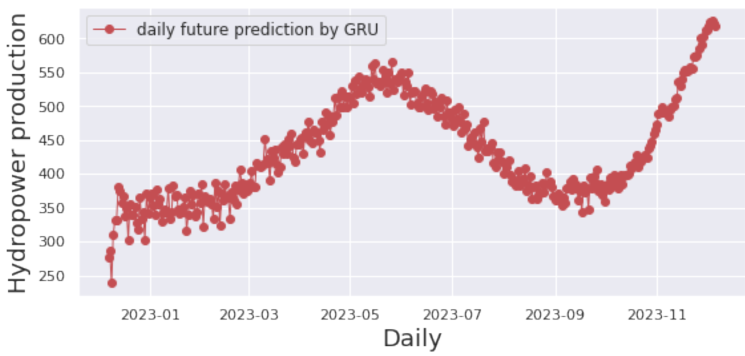

Figure 22: Daily Future Prediction.

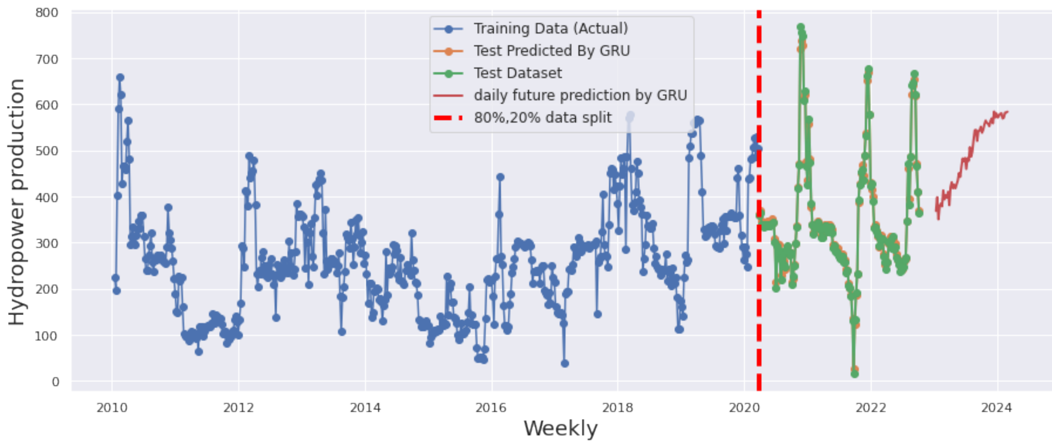

Figure 23: weekly Future Prediction.

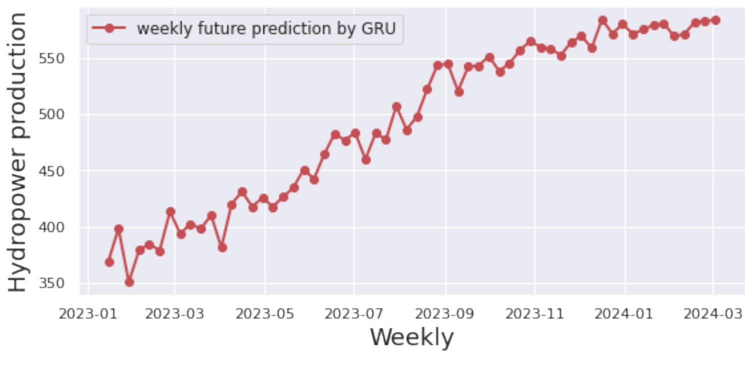

Figure 24: weekly Daily Future Prediction.

The daily one-year future hydropower production prediction graphs in Figures 21 and 22 show a sinusoidal pattern. This cyclical trend resembles historical data, suggesting seasonal variations in water flow as the primary driver of daily hydropower generation at the Koka Plant.

In contrast, the weekly prediction for the next year in Figures 23 and 24 shows a seemingly increasing trend. This contradicts the known issue of sedimentation affecting the dam's capacity. Sedimentation typically leads to a decrease in generation capacity over time. To reconcile this discrepancy and gain a clearer picture of future hydropower production, further analysis is necessary.

5 Discussion

Accurate daily and weekly hydropower production forecasts are vital in many industries, especially those that rely on real-time decision-making based on precise electric power information. Hydropower generation is a crucial component of the energy mix, and its variability can have significant impacts on the overall electric grid and related industries. In this study, the researchers investigated the accuracy of three popular machine learning algorithms - Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs), and Gated Recurrent Units (GRUs) - in predicting daily and weekly electric power generation in Koka hydropower plant, Ethiopia. These algorithms are well-suited for capturing the temporal and sequential nature of time-series data, which is essential for modeling the dynamic patterns in hydropower production.

The study provides valuable insights into how well these models can capture short-term changes in electric power generation. Understanding daily electric power generation patterns is crucial for various applications, including energy production management, weather forecasting, and the aviation industry, where daily decisions depend on accurate and reliable hydropower predictions. The generation of hydropower is highly dependent on various environmental factors, such as precipitation, maximum and minimum temperature, maximum and minimum wind speed, wind direction, relative humidity,all-sky surface shortwave downward irradiance, discharge, and hydropower production, which can fluctuate significantly on a daily and weekly basis. In both daily and weekly scenarios the GRU emerged as an outstanding performance in predicting hydropower production in Koka as compared to RNN and LSTM (Tables 11 and 12, Figure 25, Figure 26, Figure 27, and Figure 28.

The performance of the GRU (Gated Recurrent Unit) model has also been verified by different researchers worldwide. For example, compression of GRU, ARIMA regression-based model, SVM, ELM, and some Hybrid machine learning models showed outstanding performance of this model in Table 12.

The use of advanced machine learning techniques, such as RNNs, LSTMs, and

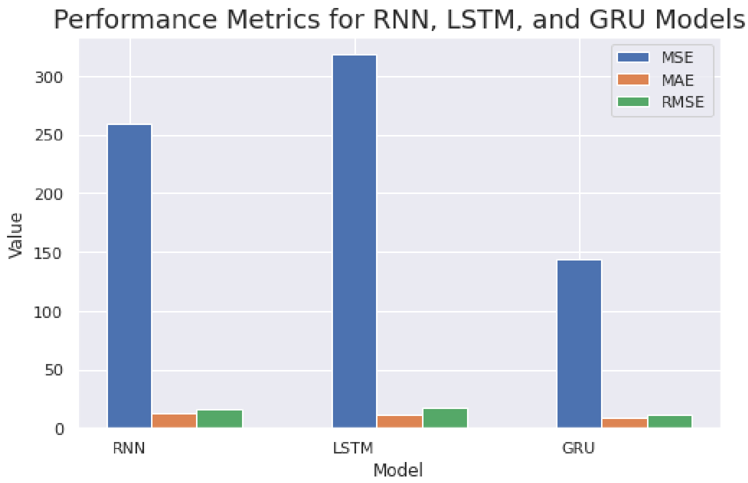

Figure 25: The bar graph shows the values for each performance metric (MSE, MAE, RMSE) for the three models side-by-side, making it easy to compare their relative performance

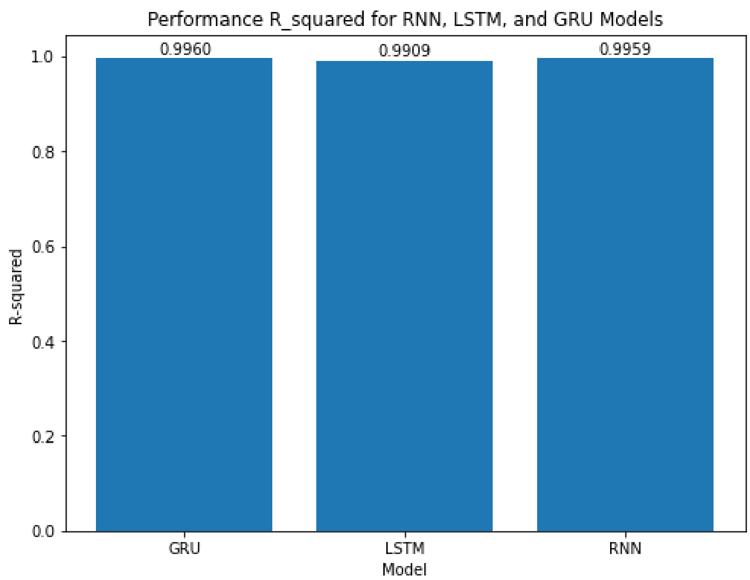

Figure 26: The bar graph shows the values for each performance metric (R-squared) for the three models side-by-side, making it easy to compare their relative performance

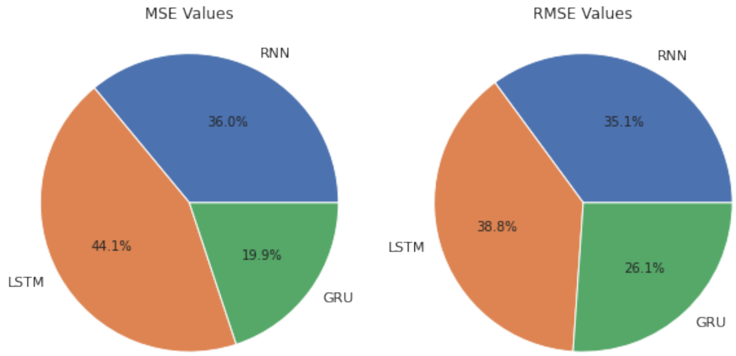

Figure 27: The circle graph focuses specifically on the MSE and RMSE Values, providing a visual representation of the proportion of variance explained by each model.

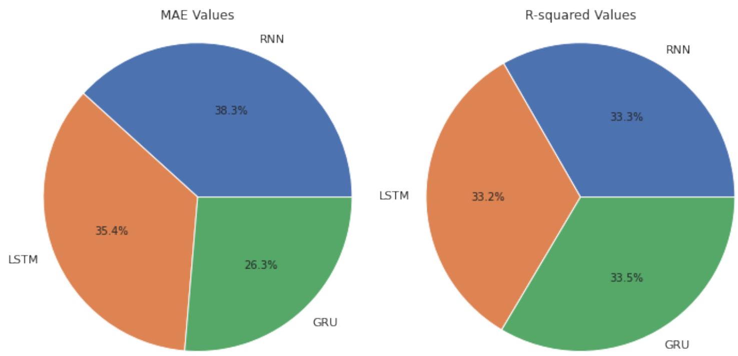

Figure 28: The Pie chart focuses specifically on the R-squared values and MAE, providing a visual representation of the proportion of variance explained by each model.

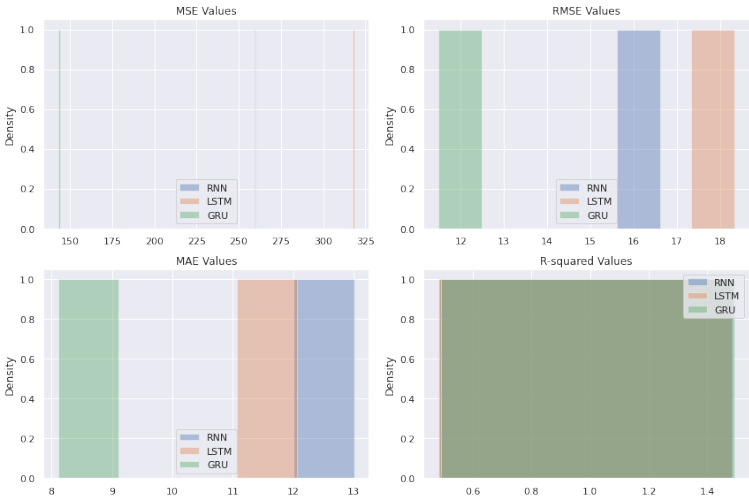

Figure 29: A histogram of the data, with the x-axis representing the values and the y-axis representing the count or frequency of each value.

Reference	Article title	Test Model Tested	best perform Model
[35]	LSTM Network-based Hybrid Model for short Term Electric load forecasting	Hybrid Model LSTM and ELM, LSTM, ELM, SVR	Hybrid model LSTM and ELM
[36]	Daily Power Generation forecasting method for Group of Small Hydropower Station	SVR, GBRT, RF, LSTM, CNN, CM-MDLN	CNN and MLP
[2]	Electric load forecasting using LSTM, GRU, and RNN	LSTM, RNN and GRU	GRU
[31]	A Comparative Study of AI Methods on Renewable Energy Prediction for Smart Grids: Case of Turkey	ANNs and LSTM, CNN and hybrid CNN-LSTM	hybrid CNN-LSTM

Reference	Article title	Test Model Tested	best perform Model
[34]	Prediction Model of Hydropower Generation and Its Economic Benefits Based on EEMD-ADAM-GRU Fusion Model	EEMD, ADAM and GRU	GRU fusion
[30]	Prediction of hydropower generation via machine learning algorithms at Three Gorges Dam, China	(ANN), (ARIMA), and SVM	ANN
[28]	Statistical model for the forecast of hydropower production in Ecuador	ARIMA, SARIMA	ARIMA
[23]	Machine learning-based small hydropower potential prediction under climate change	ANN	ANN
In This Study	-	Deep Learning Predictive Models(RNN,LSTM, and GRU)	GRU

Table 12: Related Previous Work

GRUs, has shown promising results in improving the accuracy of short-term hydropower forecasts. These models can capture the complex, nonlinear, and time-dependent patterns in hydropower generation, which can be challenging for traditional forecasting methods. By leveraging the insights gained from this study, researchers and practitioners can continue to refine and develop more accurate and reliable hydropower forecasting models. This, in turn, can lead to more efficient energy management, better integration of renewable sources, and improved decision-making across a wide range of industries that rely on precise and timely energy information.

The GRU model's exceptional R-squared value of 0.9960 for weekly forecasts, along with its remarkably low MSE, MAE, and RMSE(Table 12), demonstrates its superior ability to capture the complex, non-linear relationships and temporal dependencies inherent in hydropower systems. This is in stark contrast to the limited accuracy and predictive capabilities of the conventional approaches[2].

The key advantage of the GRU model has a simpler architecture compared to the LSTM model. This allows the GRU model to be trained and run faster. Additionally, the GRU model can learn and adapt dynamically to the unique characteristics of the hydropower system. The GRU model also has advanced feature engineering that can incorporate relevant contextual factors, which can improve the overall performance of the model. These factors contribute to the model's exceptional forecasting accuracy and its potential for broader adoption and adaptation in similar renewable energy contexts, both within Ethiopia and across the region.

By comprehensively outperforming the existing forecasting approaches, the GRU model represents a transformative leap forward in hydropower prediction, empowering the Ethiopia Electric hydropower companies to make more informed and strategic decisions to address the country's energy challenges[2].

One of the key applications of these forecasts is in energy production planning and scheduling. Energy producers need to be able to anticipate changes in hydropower generation to optimize their overall energy mix, allocate resources efficiently, and ensure a reliable supply of electricity. Inaccurate forecasts can lead to suboptimal energy production decisions, potentially resulting in higher costs, increased greenhouse gas emissions, and reduced grid resilience. Moreover, hydropower production forecasts are vital for integrating renewable energy sources, such as solar and wind, into the grid. Accurate predictions of hydropower generation can help energy system operators better balance the fluctuating output of these intermittent renewable sources, ensuring a stable and reliable electricity supply.

Beyond energy management, short-term hydropower forecasts also have important implications for other industries. In the water resources management sector, these forecasts can aid in reservoir operations, flood control, and water allocation decisions. In the aviation industry, accurate hydropower predictions can support flight planning and fuel management, as changes in energy supply can impact airport operations and flight routes.

6 Conclusion and Recommendation

6.1 Conclusion

The results of this study demonstrate the power of the deep learning approach in predicting hydropower production in the Koka hydroelectric power plant, Ethiopia. Tackling the challenge of reliable hydropower forecasting is of critical importance as energy demands continue to surge globally. The deep learning models such as RNN, LSTM, and GRU were employed on climate data, discharge, and power production data sets.

The performance metrics of this model are reported as outstanding, Achieving R-squared values consistently above 0.99, along with remarkably low mean squared errors, mean absolute errors, and root mean squared errors. Particularly, the GRU model's performance on the daily dataset showed, R-squared=0.9920, MSE=143.55, MAE=8.612, and just for RMSE=11.981 and on the weekly dataset, with an R-squared=0.9960, MSE=66.219, MAE=5.992, and RMSE=8.137. This indicates the outstanding performance of the GRU model. These accurate hydropower forecasts enable Ethhiopia's electricity sector to better manage and plan their transmission and distribution operations, ensuring reliable and uninterrupted power supply to customers.

6.2 Recommendations

The results achieved in developing reliable hydropower forecasting models for Ethiopia warrant several key recommendations to further enhance the impact and applicability of this work.

- First, Incorporate Sedimentation Data: Including sedimentation data for the region in the prediction model would allow it to account for the decreasing reservoir capacity due to sediment accumulation. This would provide a more realistic estimate of future weekly hydropower production.
- Extend the Weekly Prediction Window: Extending the prediction timeframe

for the weekly data to a range of 2 to 5 years would provide a clearer picture of the long-term trend. This could reveal the limitations of the current model in capturing the impact of sedimentation on a longer timescale

- Expanding the model evaluation to test generalizability and robustness, such as through out-of-sample forecasting, would strengthen the confidence in the model's performance.

- Investigate ensemble modeling approaches that combine the strengths of the individual LSTM, GRU, and RNN models could potentially yield even stronger predictive capabilities.

- Closely collaborating with electricity companies and grid operators in Ethiopia to align the forecasting models with their operational needs would ensure seamless integration into decision-making processes.

- Exploring the incorporation of satellite and remote sensing data as additional inputs could further enhance the models' predictive power.

- Disseminating the research findings widely through publications and conference presentations would help raise awareness and foster further collaboration on this globally important challenge of reliable hydropower forecasting.

References

[1] Ewnetu Abebe, Hailemichael Kebede, Mickus Kevin, and Zelalem Demissie. Earthquakes magnitude prediction using deep learning for the horn of africa. *Soil Dynamics and Earthquake Engineering*, 170:107913, 2023.

[2] Mobarak Abumohsen, Amani Yousef Owda, and Majdi Owda. Electrical load forecasting using lstm, gru, and rnn algorithms. *Energies*, 16(5):2283, 2023.

[3] Jesmin Akter. Bootstrapped durbin–watson test of autocorrelation for small samples. *ABC Journal of Advanced Research*, 3(2):137–142, 2014.

[4] Halit Apaydin, Hajar Feizi, Mohammad Taghi Sattari, Muslume Sevba Colak, Shahaboddin Shamshirband, and Kwok-Wing Chau. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water*, 12(5):1500, 2020.

[5] John Barnard, Robert McCulloch, and Xiao-Li Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, pages 1281–1311, 2000.

[6] Julio Barzola-Monteses, Juan Gomez-Romero, Mayken Espinoza-Andaluz, and Waldo Fajardo. Hydropower production prediction using artificial neural networks: an ecuadorian application case. *Neural Computing and Applications*, 34(16):13253–13266, 2022.

[7] Natei Ermias Benti, Mesfin Diro Chaka, and Addisu Gezahegn Semie. Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. *Sustainability*, 15(9):7087, 2023.

[8] Addisu Worku Bezabih. Evaluation of small hydropower plant at ribb irrigation dam in amhara regional state, ethiopia. *Environmental Systems Research*, 10:1–9, 2021.

[9] Mehmet Bilgili, Sinan Keiyinci, and Firat Ekinci. One-day ahead forecasting of energy production from run-of-river hydroelectric power plants with a deep learning approach. *Scientia Iranica*, 29(4), 2022.

[10] Carmen LT Borges and Roberto J Pinto. Small hydro power plants energy availability modeling for generation reliability evaluation. *IEEE Transactions on Power Systems*, 23(3):1125–1135, 2008.

[11] Paul Breeze. *Power generation technologies.* Newnes, 2019.

[12] Mesfin Diro Chaka, Addisu Gezahegn Semie, Yedilfana Setarge Mekonnen, Chernet Amente Geffe, Hailemichael Kebede, Yonas Mersha, Fikru Abiko Anose, and Natei Ermias Benti. Improving wind speed forecasting at adama wind farm ii in ethiopia through deep learning algorithms. *Case Studies in Chemical and Environmental Engineering*, 9:100594, 2024.

[13] Ernesto Chavero-Navarrete, Mario Trejo-Perea, Juan Carlos Jáuregui-Correa, Roberto Valentín Carrillo-Serrano, and José Gabriel Ríos-Moreno. Expert control systems for maximum power point tracking in a wind turbine with pmsg: State of the art. *Applied Sciences*, 9(12):2469, 2019.

[14] C Condemi, David Casillas-Perez, Loretta Mastroeni, Silvia Jiménez-Fernández, and Sancho Salcedo-Sanz. Hydro-power production capacity prediction based on machine learning regression techniques. *Knowledge-Based Systems*, 222:107012, 2021.

[15] Warwick B Elley. How in the world do students read? iea study of reading literacy. 1992.

[16] Youssef Faouzi. Hydropower forecast in ethiopia using the hbv model-a case study in omo-gibe river basin. *TVVR19/5015*, 2019.

[17] Manfred Hafner, Simone Tagliapietra, and Lucia De Strasser. *Energy in Africa: Challenges and opportunities.* Springer nature, 2018.

[18] Ashebir Dingeto Hailu. Ethiopia hydropower development and nile basin hydro politics. *AIMS Energy*, 10(1):87–101, 2022.

[19] Ashebir Dingeto Hailu and Desta Kalbessa Kumsa. Ethiopia renewable energy potentials and current state. *Aims Energy*, 9(1), 2021.

[20] Zelalem Hailu and HB Horlacher. Sustainable small hydropower development in ethiopia case study ropi hydropower plant (bilate basin, ethiopia). 2001.

[21] Mark A Hall. *Correlation-based feature selection for machine learning.* PhD thesis, The University of Waikato, 1999.

[22] IEA IEA. Renewable energy market update. *Renewable Energy Market Update*, 2022.

[23] Jaewon Jung, Heechan Han, Kyunghun Kim, and Hung Soo Kim. Machine learning-based small hydropower potential prediction under climate change. *Energies*, 14(12):3643, 2021.

[24] Bhabishya Khaniya, Chamaka Karunanayake, Miyuru B Gunathilake, and Upaka Rathnayake. Projection of future hydropower generation in samanalawewa power plant, sri lanka. *Mathematical Problems in Engineering*, 2020:1–11, 2020.

[25] Adam Krechowicz, Maria Krechowicz, and Katarzyna Poczeta. Machine learning approaches to predict electricity production from renewable energy sources. *Energies*, 15(23):9146, 2022.

[26] Jung-Pin Lai, Yu-Ming Chang, Chieh-Huang Chen, and Ping-Feng Pai. A survey of machine learning models in renewable energy predictions. *Applied Sciences*, 10(17):5975, 2020.

[27] Ayele Nigussie Legesse. Electric power demand, and the growth and transformation plan of ethiopia, 2011.

[28] Mónica Mite-León and Julio Barzola-Monteses. Statistical model for the forecast of hydropower production in ecuador. *International Journal of Renewable Energy Research*, 8(2):1130–1137, 2018.

[29] Md Hossain Alam Mondal, Abiti Getaneh Gebremeskel, Kiflom Gebrehiwot, and Claudia Ringler. *Ethiopian universal electrification development strategies.* Intl Food Policy Res Inst, 2018.

[30] Marwah Sattar Hanoon, Ali Najah Ahmed, Arif Razzaq, Atheer Y. Oudah, Ahmed Alkhayyat, Yuk Feng Huang, Pavitra kumar, and Ahmed El-Shafie. Prediction of hydropower generation via machine learning algorithms at three gorges dam, china. *Ain Shams Engineering Journal*, 14(4):101919, 2023.

[31] Derya Betul Unsal, Ahmet Aksoz, Saadin Oyucu, Josep M Guerrero, and Merve Guler. A comparative study of ai methods on renewable energy prediction for smart grids: Case of turkey. *Sustainability*, 16(7):2894, 2024.

[32] Peter M Visscher, Gibran Hemani, Anna AE Vinkhuyzen, Guo-Bo Chen, Sang Hong Lee, Naomi R Wray, Michael E Goddard, and Jian Yang. Statistical power to detect genetic (co) variance of complex traits using snp data in unrelated samples. *PLoS genetics*, 10(4):e1004269, 2014.

[33] George CS Wang and Charles K Akabay. Autocorrelation: Problems and solutions in regression modeli. *The Journal of Business Forecasting*, 13(4):18, 1994.

[34] Jiechen Wang, Zhimei Gao, and Yan Ma. Prediction model of hydropower generation and its economic benefits based on eemd-adam-gru fusion model. *Water*, 14(23):3896, 2022.

[35] Liwen Xu, Chengdong Li, Xiuying Xie, and Guiqing Zhang. Long-short-term memory network based hybrid model for short-term electrical load forecasting. *Information*, 9(7):165, 2018.

[36] Shaojun Yang, Hua Wei, Le Zhang, and Shengchao Qin. 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, 2021.