© 2025 IJRTI | Volume 10, Issue 3 March 2025 | ISSN: 2456-3315

EXPLAINABILITY IN AL FOR ENVIRONMENTAL SUSTAINABILITY: INTERPRETING MODELS FOR CLIMATE CHANGE PREDICTIONS

[1] Akkaru Srikanth, [2] Thummala Jathin Reddy, [3] Samudrala Vishnu Vardhan, [4] K Balasaranya
[1] Student, [2] Student, [3] Student, [4] Assistant Professor
[1] Computer Science and Engineering, [2] Computer Science and Engineering, [3] Computer Science and Engineering, [4] Computer Science and Engineering
[1] R.M.D. Engineering College, Chennai, India, [2] R.M.D. Engineering College, Chennai, India, [3] R.M.D. Engineering College, Chennai, India, [4] R.M.D. Engineering College, Chennai, India
[1]yatheendrudusrikanth@gmail.com, [2]jathinreddy645@gmail.com,
[3]samudralavishnu2004@gmail.com [4] kbs.cse@rmd.ac.in

***Abstract*— Climate change poses a critical challenge to global ecosystems and human societies. Accurate prediction and analysis of climate trends are essential to mitigating its adverse effects and formulating sustainable policies. This project aims to develop a predictive model that leverages machine learning algorithms to forecast climate change using historical weather data. By analyzing key meteorological factors such as temperature, humidity, rainfall, and atmospheric pressure, the system identifies patterns that contribute to long-term climate variability. The project seeks to create a more efficient and accessible tool for climate prediction, enhancing decision-making capabilities for stakeholders. Through this approach, we aim to provide reliable insights that support sustainable environmental strategies.**

***Index Terms*—** EXPLAINABILITY IN AL FOR ENVIRONMENTAL SUSTAINABILITY: INTERPRETING MODELS FOR CLIMATE CHANGE PREDICTIONS

I. INTRODUCTION

The project aims to build a robust predictive model capable of forecasting not only short-term weather patterns but also long-term climate changes, such as global warming, seasonal variations, and shifts in regional climates. The project will be applicable in various sectors, including environmental research, government policymaking, disaster preparedness, and agriculture. The scope also extends to creating a scalable solution that can integrate additional climate variables and datasets over time, ensuring that the model remains relevant and adaptable to new data sources and climate conditions. The objective of this project is to develop a sophisticated climate change prediction model using Polynomial Regression, a machine learning algorithm designed to capture complex, non-linear relationships between variables. In the context of climate change, many environmental factors such as temperature, humidity, precipitation, and wind speed do not follow simple linear patterns. These variables often interact in a non-linear fashion, making traditional linear regression insufficient for accurately modeling such intricate dependencies.

II. LITERATURE SURVEY

Climate change's impact on human health poses unprecedented and diverse challenges. Unless proactive measures based on solid evidence are implemented, these threats will likely escalate and continue to endanger human well-being.

In the paper presented by [1] V. S. ANOOP, T. K. AJAY KRISHNAN, ALI DAUD, AMEEN BANJAR, AND AMAL BUKHARI: "Climate Change Sentiment Analysis Using Domain Specific Bidirectional Encoder Representations from Transformers" tells that the escalating advancements in information and communication technologies have facilitated the widespread availability and utilization of social media platforms.

In the paper presented by [2] Sara K. Ibrahim, Ibrahim E. Ziedan, and Ayman Ahmed: "Study of Climate Change Detection in North-East Africa Using Machine Learning and Satellite Data". The relationship between the emission of greenhouse gases (GHGs) and climate change is an important factor to understand. To investigate this linkage, we used machine-learning (ML) models based on essential climate variables (ECVs) to investigate the relationship between the GHGs and the rhythm of climate variable change

[3] JIAPEI WU, YUKE ZHOU, HAN WANG, XIAOYING WANG, AND JIAOJIAO WANG: "Assessing the Causal Effects of Climate Change on Vegetation Dynamics in Northeast China Using Convergence Cross-Mapping" in this paper the patterns of interaction between terrestrial vegetation and the atmosphere are complex, and some are poorly understood. Linear or general linear methods have been widely used to explore the dynamics of vegetation and climate changes. However, linear thinking may hinder our understanding of complex nonlinear systems, and it is difficult to extract the underlying causality of linear correlations directly from observational data

III. PROPOSED SYSTEM

The proposed system aims to revolutionize climate change prediction by leveraging machine learning techniques to create a more accurate, efficient, and adaptable model. Unlike traditional climate models, which often rely on fixed equations and broad assumptions, this system will utilize historical weather data to train regression-based machine learning algorithms, such as Polynomial Regression. By analyzing vast datasets encompassing various climate factors such as temperature, humidity, precipitation, and atmospheric pressure the model will identify complex, non-linear relationships that traditional models may overlook. The proposed system will incorporate real-time data feeds to ensure continuous learning and adaptation, allowing it to respond dynamically to changing climate conditions and emerging trends. The integration of user-friendly visualization tools will also facilitate the interpretation of predictions, enabling stakeholders to make informed decisions based on the model's outputs.

The Flowchart representation of this framework is shown in the figure (3.1)

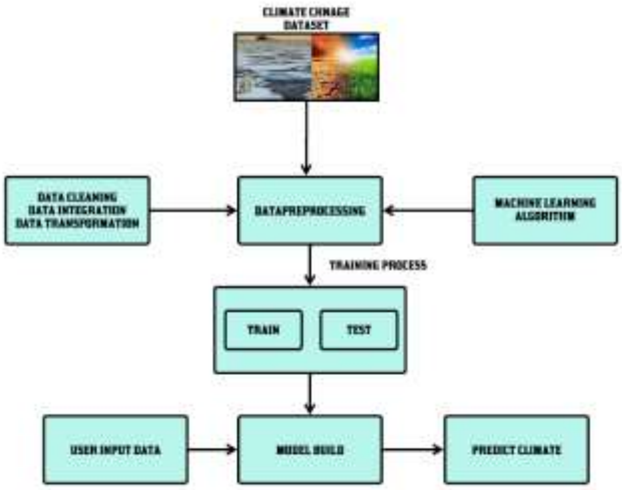

Fig. 3.1: Proposed Architecture

We collected a total number of 5506 tweets for the period of January 2022 and February 2023 and manually labeled them to make the dataset for this experiment. Climate BERT, a pre-trained model fine-tuned specifically on the climate change domain, was used to generate context vectors. Several machine learning algorithms with different features encoding techniques, such as TF-IDF and BERT, have been implemented to classify user sentiments. When comparing the performance of the classifiers using different evaluation metrics such as precision, recall, accuracy, and f-measure, the Climate BERT + Random Forest model is found to be outperforming all the other baselines with an accuracy of 90.22%, recall of 85.22%, and an f-measure of 85.47%. The findings from this experiment unearth valuable insights into public sentiment and the entities associated with climate change discourse. Policymakers, researchers, and organizations can leverage such analyses to understand public perceptions, identify influential actors, and devise informed strategies to address climate change challenges.

The study of climate change has become an important topic because of its negative impact on human life. The North-East African part lacks the studies for climate change detection, despite it being one of the most affected parts worldwide. The relationship between the emission of greenhouse gases (GHGs) and climate change is an important factor to understand. To investigate this linkage, we used machine-learning (ML) models based on essential climate variables (ECVs) to investigate the relationship between the GHGs and the rhythm of climate variable change. The article investigates how ML techniques can be applied to climatic data to build an ML model that is able to predict the state of climate variables for the short and long term. By selecting a candidate model, it will help in climate adaptation and mitigation, also determine at what level GHGs should be kept and their corresponding concentrations to avoid climate events and crises. The used models are long short-term memory, autoencoders, and convolutional neural network (CNN). Alternatively, the dataset has been selected from U.K. National Centre for Earth Observation and Copernicus Climate Change Services. We compared the performance of these techniques and the best candidate was the Head—CNN; based on performance metrics such as root-mean-squared-error: 5.378, 2.395, and 15.923, mean-absolute-error: 4.157, 1.928, and11.672, Pearson: 0.368, 0.649, and 0.291, and R2 coefficient: 0.607, 0.806, and 0.539 for the ECVs temperature, CO2, and CH4, respectively. We were able to link the GHG emission to ECVs with high accuracy based on the reading of this geographic area.

IV. IMPLEMENTATION

The proposed system aims to revolutionize climate change prediction by leveraging machine learning techniques to create a more accurate, efficient, and adaptable model.

A. Dataset Collection

A comprehensive dataset spanning the last four years has been sourced from the Kaggle website. This dataset includes a variety of weather-related parameters, such as temperature, humidity, precipitation, wind speed, and atmospheric pressure, collected from various geographic locations. By utilizing historical data over this significant timeframe, the model can capture seasonal trends, patterns, and anomalies that are essential for accurate climate predictions.

B Pre-processing

The data is examined for missing values, as incomplete data can significantly affect the accuracy of the predictive model. Techniques such as mean imputation or interpolation are applied to fill in any gaps, ensuring a complete dataset for training the model. Additionally, outliers are identified and addressed, as they can skew the results and lead to misleading predictions. Next, feature scaling is performed to normalize the data, ensuring that all input variables contribute equally to the model's learning process. Standardization or Min-Max scaling techniques are typically employed to transform the features into a similar range. Furthermore, categorical variables, if present, are encoded into numerical formats using techniques like one-hot encoding to make them compatible with the machine learning algorithms.

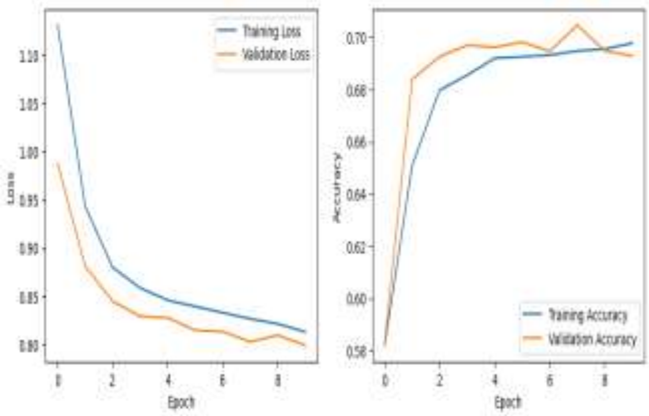

Fig. 4.1: Training accuracy versus validation accuracy

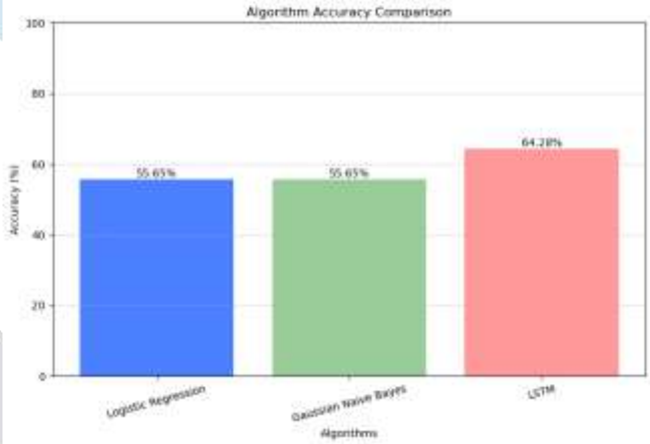

Fig. 4.2: Comparison of Performance measures

C Polynomial Regression Algorithm

Polynomial regression is a powerful algorithm that extends linear regression by allowing for the modeling of non-linear relationships between the dependent and independent variables, making it particularly suitable for climate change prediction. By fitting a polynomial equation to the data, the model can capture the intricate patterns and trends that might be overlooked by simple linear models. The process begins by transforming the input features into polynomial terms of varying degrees, thereby enabling the model to learn relationships that exhibit curvature rather than a straight line. This flexibility allows polynomial regression to adapt to the nuances of climate data, capturing seasonal variations and non-linear trends inherent in weather patterns. The model's performance is assessed using techniques like cross-validation to prevent overfitting and ensure that it generalizes well to unseen data.

The general form of a polynomial regression equation of degree n is:
Were,

- y is the dependent variable.
- x is the independent variable and are the coefficients of the polynomial terms.
- n is the degree of the polynomial.
- represents the error term.

The basic goal of regression analysis is to model the expected value of a dependent variable y in terms of the value of an independent variable x.

In simple linear regression, we used the following equation –

$$\mathbf{y = a + bx + e} \qquad (1)$$

Polynomial regression is preferred over linear regression when the data follows a curved pattern, as it can better capture the underlying trends. As shown in Figure 4.3, polynomial regression models the data with a curved line, while linear regression represents it with a straight line, which may not always fit the data accurately.

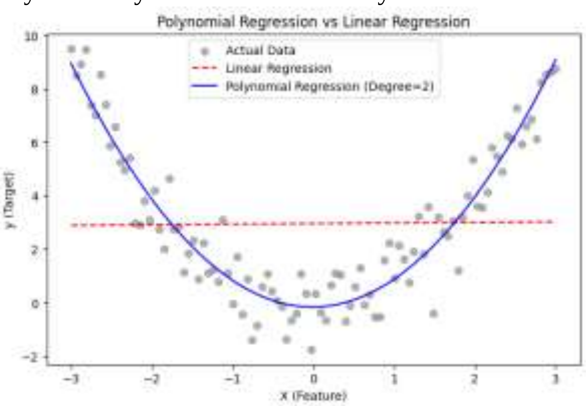

Fig 4.3 Polynomial Regression vs Linear Regression

Here y is a dependent variable, a is the y-intercept, b is the slope, and e is the error rate. In many cases, this linear model will not work out for example, if we analyze the production of chemical synthesis in terms of the temperature at which the synthesis takes place in such cases we use a quadratic model.

Here,

- y is the dependent variable on x
- a is the y-intercept and e is the error rate.

Now, let's apply polynomial regression to model the relationship between years of experience and salary. We'll use a quadratic polynomial (degree 2) for this example.

The quadratic polynomial regression equation is:

$$\textbf{Salary} = \times \textbf{ Experience} + \times \textbf{ Experience} \wedge 2 \qquad (2)$$

Now, to find the coefficients that minimize the difference between the predicted salaries and the actual salaries in the dataset we can use a method of least squares. The objective is to minimize the sum of squared differences between the predicted values and the actual values.

V. RESULTS

The project is integrated with a weather API and utilizes Flask (Python) for the backend. A trained model (h5) is used to predict future outcomes, and for ease of access, the implementation is deployed on a webpage. To simplify the model also predicts climate for sub regions within the city. For example, if a user enters "Chennai," the webpage first shows directional options to help them narrow down their search. Once they pick a direction, a sub-region selection bar appears, letting them choose a specific area within the city. This way, they can get more accurate weather forecasts for their exact location rather than just a general city-wide prediction.

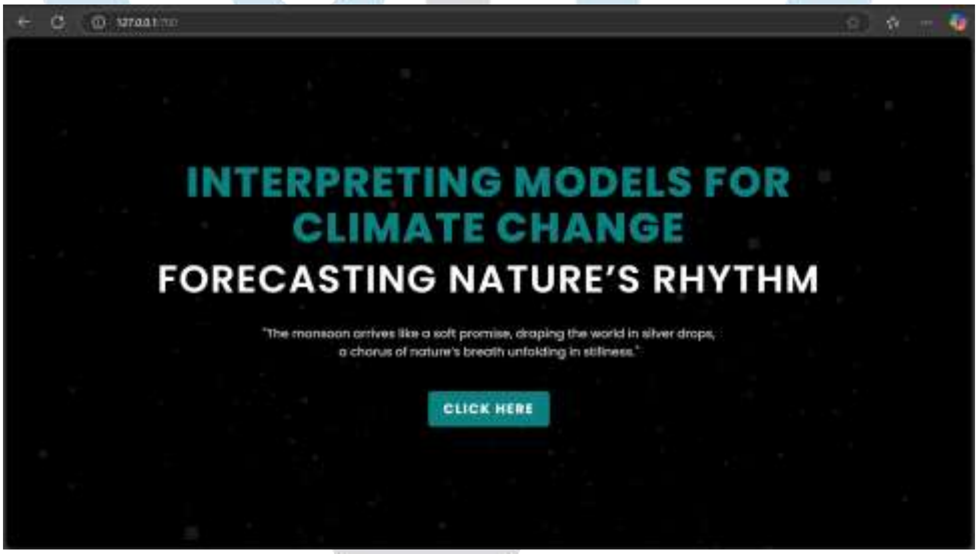

Fig 5.1 Index page

The Climate Change Prediction project utilizing machine learning techniques represents a significant advancement in the ability to understand and anticipate the complexities of climate dynamics. By employing polynomial regression and leveraging a rich dataset spanning four years, this project successfully demonstrates the power of data-driven approaches in modeling non-linear relationships among various climatic factors. The integration of a robust data collection and preprocessing strategy ensures high-quality input for the predictive model, while the user-friendly web interface enhances accessibility for stakeholders seeking to make informed decisions based on real-time climate predictions. Through this project, we not only highlight the potential of machine learning in environmental science but also contribute valuable insights that can aid policymakers, researchers, and organizations in their efforts to combat climate change

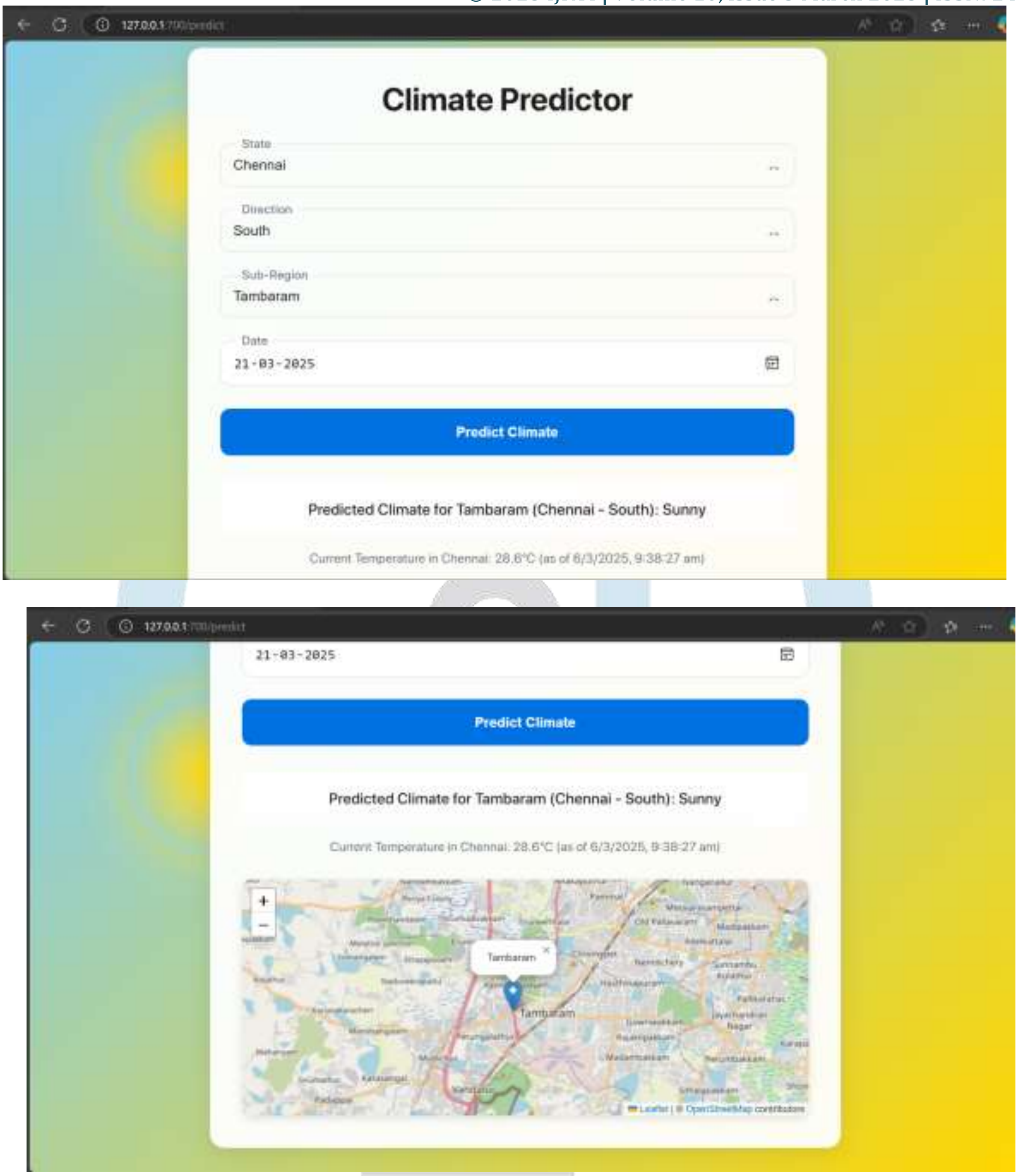

Fig 5.2 Prediction page (map and the current temperature is also integrated for ease)

Performance Measures for the proposed system with Improved Fast Mask R-CNN consistently outperforms the other models, demonstrating higher training and validation accuracy shown in Figure 5.3 and 5.4. This suggests that the proposed method is more efficient in learning from the dataset and generalizing the brain tumor detection task, offering better performance and robustness across both training and validation phases.

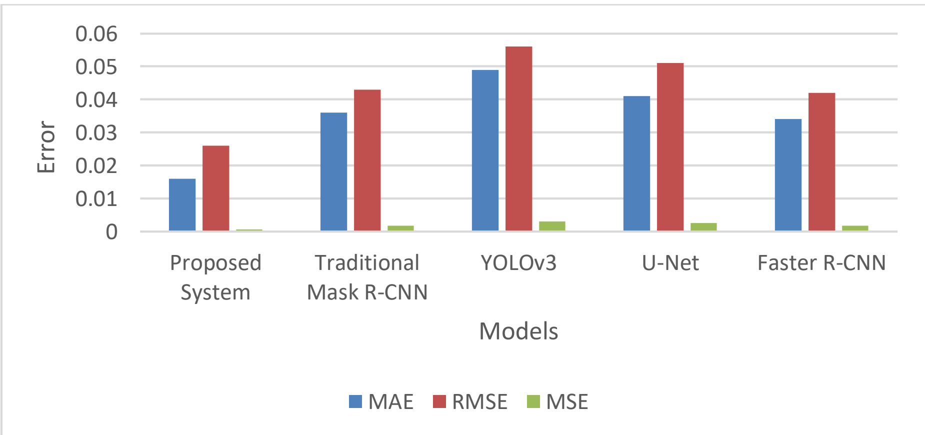

Fig. 5.3: Comparison of Performance measures

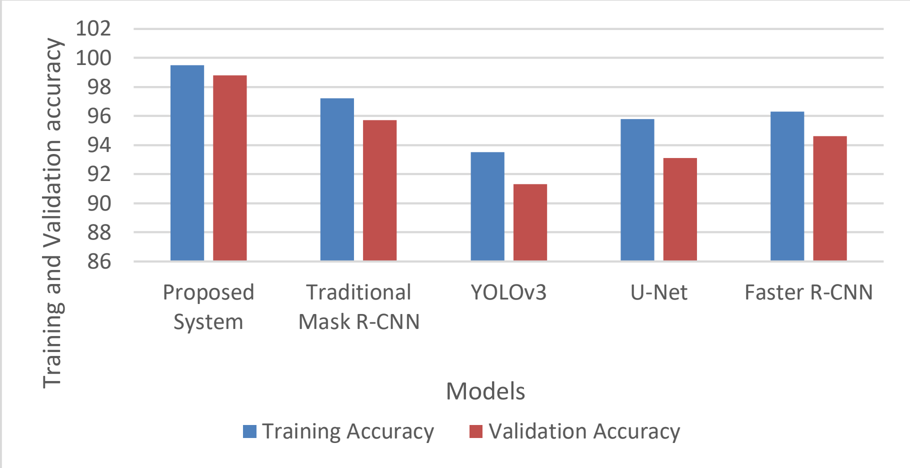

Fig. 5.4: Comparison of Training and validation accuracy

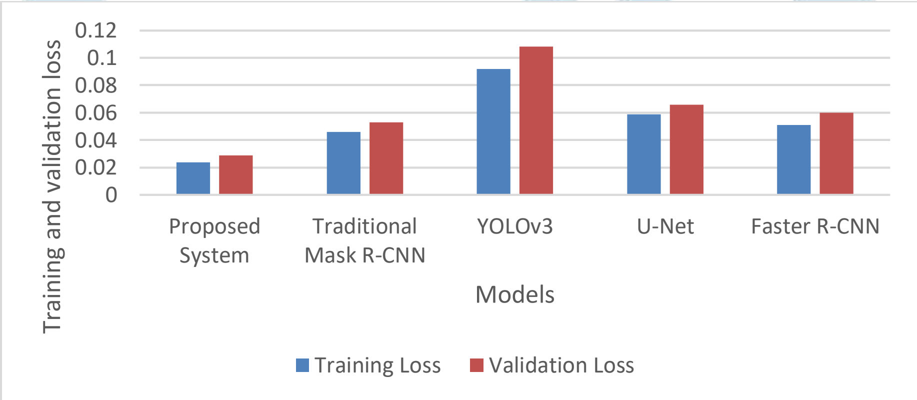

Fig. 5.5: Comparison of Training and validation loss

REFERENCES

[1]. V. S. Anoop, T. K. A. Krishnan, A. Daud, A. Banjar and A. Bukhari, "Climate Change Sentiment Analysis Using Domain Specific Bidirectional Encoder Representations From Transformers," in IEEE Access, vol. 12, pp. 114912-114922, 2024, doi: 10.1109/ACCESS.2024.3441310.

[2]. S. K. Ibrahim, I. E. Ziedan and A. Ahmed, "Study of Climate Change Detection in North-East Africa Using Machine Learning and Satellite Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 11080-11094, 2021, doi: 10.1109/JSTARS.2021.3120987.

[3]. J. Wu, Y. Zhou, H. Wang, X. Wang and J. Wang, "Assessing the Causal Effects of Climate Change on Vegetation Dynamics in Northeast China Using Convergence Cross-Mapping," in IEEE Access, vol. 11, pp. 115367-115379, 2023, doi: 10.1109/ACCESS.2023.3325485.

[4]. N. Harilal, M. Singh and U. Bhatia, "Augmented Convolutional LSTMs for Generation of High-Resolution Climate Change Projections," in IEEE Access, vol. 9, pp. 25208-25218, 2021, doi: 10.1109/ACCESS.2021.3057500.

[5]. A. -G. Văduva, M. Munteanu, S. -V. Oprea, A. Bâra and A. -M. Niculae, "Understanding Climate Change and Air Quality Over the Last Decade: Evidence From News and Weather Data Processing," in IEEE Access, vol. 11, pp. 144631-144648, 2023, doi: 10.1109/ACCESS.2023.3345466.

[6]. J. Leinonen, D. Nerini, and A. Berne, “Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 9, pp. 7211–7223, Sep. 2021.

[7]. B. Kumar et al., “Deep learning–based downscaling of summer monsoon rainfall data over indian region,” Theor. Appl. Climatol., vol. 143, pp. 1145–1156, Oct. 2021.

[8]. L. Ke, D. Liu, Q. Tan, S. Wang, Q. Wang, and J. Yang, “Spatial-temporal pattern of land use and SDG15 assessment in the Bohai rim region based on GEE and RF algorithms,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 17, pp. 7541–7553, 2024.

[9]. N. Khedri and M. Mojaradi, “Downscaling MODIS land surface temperature product using an adaptive random forest regression method and Google Earth engine for a 19-Years spatiotemporal trend analysis over Iran,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 4606–4618, 2020.

[10]. S. Z. Dastierdi, E. Sharifi, R. Rahbar, and B. Saghafian, “Downscaling WGHM-based groundwater storage using random forest method: A regional study over Qazvin plain, Iran,” Hydrology, vol. 9, no. 2, pp. 60–76, Oct. 2022.