HARNESSING AI DRIVEN PREDICTION OF TOTAL ELECTRON CONTENT IN IONOSPHERE

A PROJECT REPORT

Submitted by

AMBIKA. K

JANANI. D

RUPA SRI VARSINI. R

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

SARANATHAN COLLEGE OF ENGINEERING
(An Autonomous Institution)

Tiruchirappalli-620012

ANNA UNIVERSITY: CHENNAI 600 025

APRIL 2025

SARANATHAN COLLEGE OF ENGINEERING

(An Autonomous Institution)

Tiruchirappalli-620012

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **TOTAL ELECTRON CONTENT PREDICTION IN IONOSPHERE USING MACHINE LEARNING TECHNIQUES** is the Bonafide work of **AMBIKA. K, JANANI. D, RUPA SRI VARSINI. R** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr. S Ravimaran, M.E., Ph.D.,	**Mrs. A Sridevi, M.E.,**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Assistant Professor
Artificial Intelligence and Data Science	Artificial Intelligence and Data Science
Saranathan College of Engineering	Saranathan College of Engineering
Tiruchirappalli –620012.	Tiruchirappalli–620012.

Submitted for the project viva-voce examination held on_____.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

The word "thanks," though apparently short, carries deep eloquence that is enhanced when expressed with true heartfelt sincerity. At this point, we would like to extend our heartfelt thanks to all the people who helped in the process of completing this important assignment.

Firstly, we want to thank the "Almighty" for granting us the "Will and Determination" to doggedly strive for our aim.

We wish to express our gratitude to our college Secretary **Shri. S. Ravindran** and our Principal **Dr. D. Valavan M.Tech., Ph.D.,** for their support in the successful completion of the project work.

With a sincere feeling of gratitude, we convey our deepest gratitude to our Head **Dr. S Ravimaran M.E., Ph.D.,** Department of Artificial Intelligence and Data Science, for his inspiration to finish the work successfully.

We take great pleasure in expressing our sincere appreciation to our project guide and the project coordinator **Mrs. A Sridevi, M.E.,** members of the project review committee and all faculty members of the department of Artificial Intelligence and Data Science for their helpful guidance, excellent support and encouragement during our project work.

We would like to extend our heartfelt thanks to our loving parents, siblings, and friends for their ongoing support.

ABSTRACT

Total Electron Content (TEC) variations in the ionosphere are crucial for understanding space weather interactions and their impact on communication and navigation systems. Plasma bubbles, observed as TEC depletion, result from large-scale irregularities in the equatorial ionosphere. Traditional methods for detecting and predicting these variations rely on empirical models with dataset-specific assumptions, limiting their generalization capabilities. In this study, we develop a machine learning-based framework to predict TEC variations during different space-weather and geophysical events, including solar flares, geomagnetic storms, and earthquakes. We compare the performance of two machine learning models Random Forest Regressor and XGBoost Regressor in predicting TEC depletion. The models are trained on verified TEC depletion events using historical and real-time data, ensuring robustness in capturing ionospheric disturbances. Our results demonstrate that both models effectively identify TEC anomalies and forecast ionospheric variations, with one model outperforming the other in specific scenarios. The study highlights the potential of ML-based TEC prediction for improving early-warning systems and mitigating the adverse effects of ionospheric disturbances in the Indian low-latitude region.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

Abbreviation	Description
AI	Artificial Intelligence
AP	Planetary Geomagnetic Activity Index
ARMA	Auto Regressive Moving Average
Bi-LSTM	Bidirectional Long Short-Term Memory
DL	Deep Learning
DST	Disturbance storm time index
EIA	Equatorial Ionization Anomaly
EPB_s	Equatorial Plasma Bubbles
F10.7	Solar Flux
GB	Gigabyte
GBR	Gradient Boost Regression
GEONET	GNSS Earth Observation Network System
GNSS	Global Navigation Satellite System
GPS- VTEC	Global Positioning System -Vertical total electron content
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
IDE	Integrated Development Environment
IGS	International GNSS Service
IISC	Indian Institute of Science
IMF	Interplanetary Magnetic Field
IRI	International Reference Ionosphere
KP	Planetary K-index
L1	Lasso Regularization

L2	Ridge Regularization
Light GBM	Light Gradient Boosting Machine
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
Mg II index	Magnesium II Index
ML	Machine Learning
MSD	Mean Square Deviation
MSLE	Mean Squared Logarithmic Error
NN	Neural Networks
PCA	Principal Component Analysis
R^2	R-squared Score
RF	Random Forest
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
SMAPE	Symmetric mean absolute percentage error
SSD	Solid-State Drive
SSN	sunspot number
TEC	Total Electron Content
UTC	Coordinated Universal Time
XGB	Extreme Gradient Boosting
XGBoost Regressor	Extreme Gradient Boosting Regressor

CHAPTER 1

INTRODUCTION

1.1 CONTEXT AND MOTIVATION

In recent years, the rising need for accurate space weather forecasting has gained significant attention, particularly in understanding and countering the disruptive effects of solar activity on Earth’s ionosphere. Total Electron Content (TEC) is a key metric that quantifies electron density in the ionosphere and is vital for satellite communication and navigation systems. Fluctuations in TEC can cause plasma bubbles that disrupt navigation satellite systems, making accurate predictions crucial for mitigating errors. Solar flares, which are intense bursts of radiation resulting from the release of magnetic energy stored in the sun's atmosphere, can significantly impact the ionosphere by increasing ionization levels, leading to abrupt changes in TEC. These flares can produce high-energy particles and electromagnetic radiation that penetrate the Earth's atmosphere, causing rapid increases in electron density and resulting in fluctuations in TEC. Similarly, solar storms, which are large-scale disturbances in the solar wind and magnetic field, can enhance the ionospheric response, causing further fluctuations in TEC. The interaction between the solar wind and the Earth's magnetic field can lead to geomagnetic storms, which can further disrupt the ionosphere and affect satellite signals. Earthquakes, while primarily geological events, have been observed to correlate with ionospheric anomalies, potentially due to the release of gases and electromagnetic signals prior to seismic activity, which may also influence TEC variations. This research focuses on predicting TEC fluctuations caused by solar flares, solar storms, and earthquakes in the year 2024, aiming to analyze the impact of significant space weather events, including solar storms occurring from May 6 to May 17, 2024, and from October 6 to October 15, 2024. Additionally, key solar flares analyzed include an X6.7 flare on February 22,

2024, and an X8.79 flare on May 14, 2024. Furthermore, the study investigates the impact of earthquakes that occurred on January 1, 2024, and October 26, 2013. To predict TEC variations, this research employs the Random Forest Regressor and XGBoost Regressor algorithms, which are well-suited for handling complex, non-linear relationships in the data. The models are evaluated using performance metrics such as R-squared (R2), Mean Squared Logarithmic Error (MSLE), and Symmetric Mean Absolute Percentage Error (SMAPE). The predicted values obtained from the Random Forest Regressor, XGBoost Regressor, and IRI 2020 models are compared with the actual TEC values to assess their accuracy. The training dataset, covering the period between 2023 and 2024, consists of real observations collected from IONOLAB, IRI 2020, and OmniWeb. The models utilize key space weather parameters, including sunspot number (SSN), planetary Ap index, geomagnetic Kp index, and F10.7 solar flux, which are critical indicators of solar activity and its potential impact on the ionosphere. The sunspot number (SSN) serves as a measure of solar activity, with higher numbers indicating increased solar radiation and potential for ionospheric disturbances. The planetary Ap index quantifies geomagnetic activity, reflecting the influence of solar wind on the Earth's magnetic field, while the Kp index provides a global measure of geomagnetic activity, indicating the strength of disturbances in the ionosphere. The F10.7 solar flux, a measure of solar radio emissions at a wavelength of 10.7 cm, is directly related to solar activity and is a reliable predictor of ionospheric conditions. By leveraging these datasets and models, this study aims to enhance the accuracy of TEC forecasting, ultimately contributing to the reliability of satellite-based communication and navigation systems under extreme space weather conditions. Additionally, the Random Forest Regressor and XGBoost Regressor models are implemented to analyze feature importance and optimize TEC prediction accuracy. These models effectively handle non-linearity in TEC variations and improve forecasting performance under different space weather conditions. In addition to predictive

modeling, this study explores the spatial and temporal variability of TEC during extreme space weather events. By examining regional TEC deviations, the research aims to identify patterns in ionospheric disturbances that could improve early warning systems for navigation and communication disruptions. Furthermore, the study evaluates the adaptability of machine learning models under varying geomagnetic conditions, enhancing their applicability for real-time forecasting. The insights gained from this research contribute to the broader understanding of ionospheric dynamics, ultimately supporting the development of more resilient global positioning and communication networks. This study also looks at how TEC changes in different seasons and locations around the world. By studying these patterns, we can understand how space weather affects different regions. The research also checks how well space weather indicators, like sunspots and geomagnetic activity, help predict TEC changes. The goal is to make forecasting more reliable, so satellite navigation and communication systems can work better even during strong space weather events.

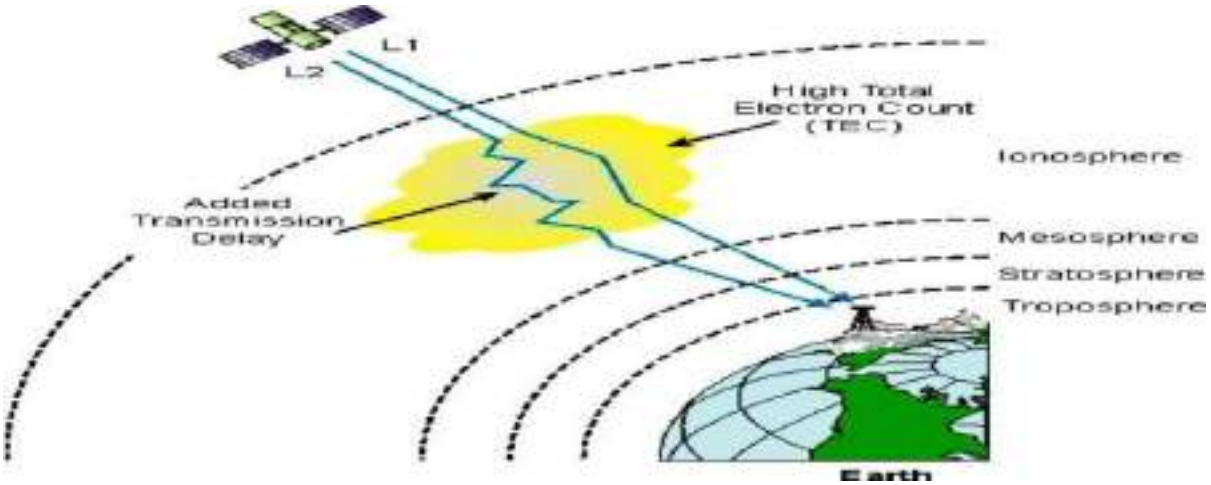

Figure 1.1: Effect of High Total Electron Content (TEC) on Satellite Signal Transmission

1.2 PROBLEM STATEMENT

Fluctuations in the ionospheric Total Electron Content (TEC) caused by extreme space weather and seismic events disrupt satellite communication and navigation systems. Existing models often lack the precision required for real-time forecasting under complex environmental conditions. There is a critical need for advanced predictive tools that can accurately model TEC variability using reliable space weather indicators and machine learning techniques

1.3 AIM OF THE PROJECT

- The aim of this study is to develop and evaluate machine learning-based models for accurate prediction of ionospheric Total Electron Content (TEC) fluctuations.
- It focuses on understanding the influence of extreme space weather events—such as solar flares, solar storms, and earthquakes—on TEC variability.
- The research aims to improve forecasting accuracy using Random Forest and XGBoost Regressor models trained on real observational data.
- By analyzing the contribution of key solar and geomagnetic parameters (SSN, Kp, Ap, F10.7), the study seeks to optimize prediction performance.
- Ultimately, the goal is to enhance the reliability of satellite-based communication and navigation systems under dynamic ionospheric conditions.

1.4 OBJECTIVES

- To analyze the impact of solar flares, solar storms, and earthquakes on TEC variations.
- To predict TEC using Random Forest and XGBoost regressors based on key space weather indicators (SSN, Kp, Ap, F10.7).

- To compare model predictions with actual TEC and IRI 2020 values using R^2, MSLE, and SMAPE metrics.
- To evaluate feature importance and model adaptability under varying geomagnetic conditions.
- To explore spatial and seasonal TEC variability during extreme space weather events.

1.5 DATA AND METHODOLOGY

The datasets used in this study were sourced from IONO Lab, IRI 2020, and OMNIWeb, providing a reliable foundation for training and testing the prediction models. The training data spans from 2023 to 2024, covering a range of solar, geomagnetic, and ionospheric conditions. Key parameters included Sunspot Number (SSN), F10.7 cm Solar Flux, Ap index, Kp index, and True TEC values — all essential for capturing the dynamic behavior of the ionosphere. Additionally, we incorporated data from notable space weather events in 2024, such as solar storms between May 6 to May 17 and October 6 to October 15, as well as solar flares like X6.7 on February 22 and X8.79 on May 14. Earthquake data from January 1, 2024, and October 26, 2013, was also included to observe potential ionospheric disturbances associated with seismic activity, providing a broader understanding of TEC variations beyond solar influences. For preprocessing, missing data was handled through linear interpolation to maintain time series continuity. Feature engineering combined key parameters with time-based markers (date, time) to help models learn event driven patterns more effectively. We implemented two machine learning models — Random Forest Regressor and XGBoost Regressor — for TEC prediction. Random Forest Regressor, an ensemble learning method, reduces overfitting by averaging multiple decision trees. We tuned hyperparameters like the number of trees (n_estimators) and maximum depth (max_depth) using Grid Search Cross-Validation. XGBoost Regressor, a gradient boosting algorithm known for speed

and performance, was optimized with hyperparameters such as learning rate, maximum depth, and regularization terms to improve prediction accuracy. The IRI 2020 model served as a baseline for comparison to evaluate the improvements made by our machine learning approaches. The training period covered data from 2023 to 2024, Trained the model before the date of Prediction. This setup allowed the models to learn from diverse patterns in ionospheric behavior while being tested on unseen data, ensuring that they could generalize effectively to new solar flare events. Special attention was given to periods of intense solar activity, particularly May 2024, to assess model performance under extreme space weather conditions. By integrating multiple datasets, preprocessing steps, and advanced machine learning techniques, this methodology provides a strong foundation for predicting TEC fluctuations during solar flares, storms, and seismic events, contributing to a better understanding of ionospheric behavior during such disturbances.

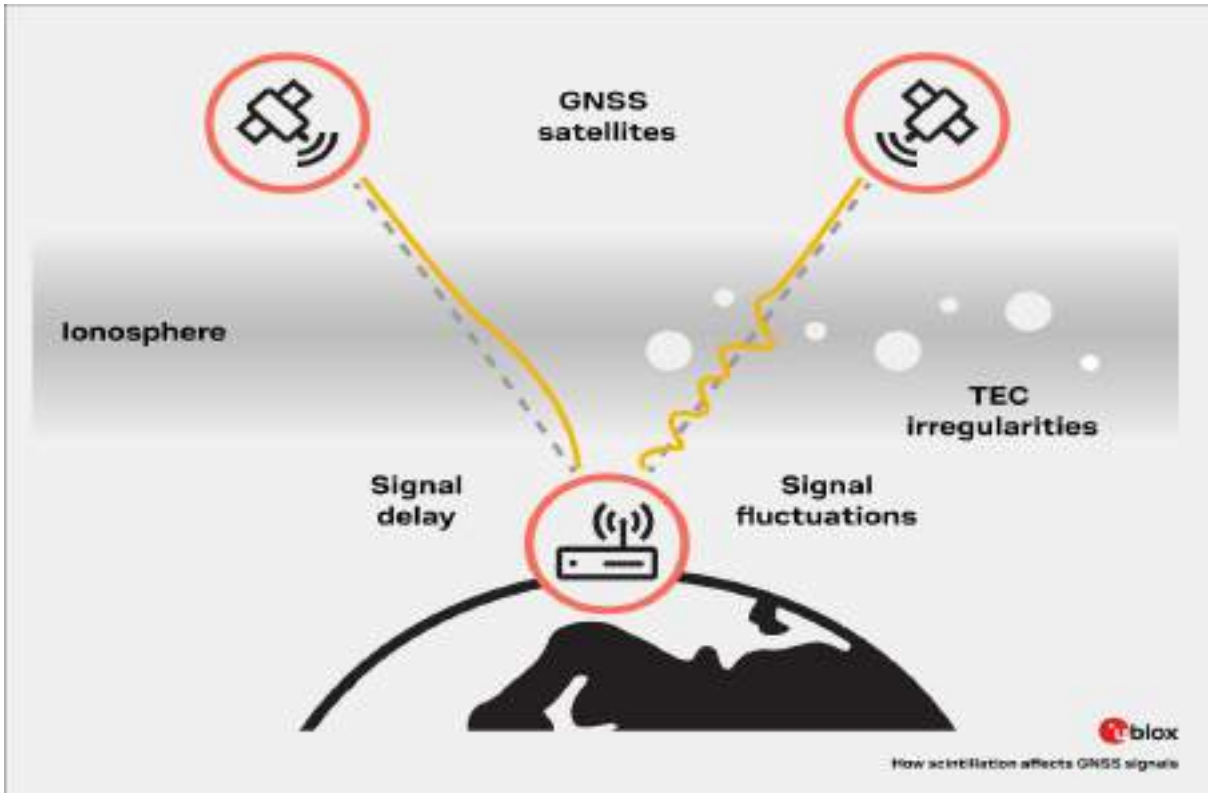

Figure 1.2: Impact of TEC Irregularities on GNSS Signal Delay and Fluctuations

CHAPTER 2

LITERATURE SURVEY

2.1 MACHINE LEARNING APPROACH FOR DETECTION OF PLASMA DEPLETIONS FROM TEC

YEAR: 2024

AUTHORS: Chandan Kapil, Gopi K. Seemala

Abstract

The paper discusses the concern of ionospheric delay for trans-ionospheric radio communication, particularly for navigation systems relying on satellite signals.It highlights that most ionospheric delays are estimated using dual frequency GNSS receivers, except during irregularities and equatorial plasma bubbles.Plasma bubbles are identified as decreases in total electron content (TEC) due to large-scale irregularities in the equatorial ionosphere.The study explores various machine learning (ML) techniques for detecting TEC depletions, with the Random Forest Regressor Method (RFM) showing superior performance.The RFM achieved a training accuracy of 97.6% and a positive predictive value (PPV) of 96.8%.The results indicate that ML techniques can surpass traditional mathematical algorithms in detecting TEC depletions, with potential for further advancements.

Advantages

- High Accuracy: The Random Forest Regressor Method achieved a training accuracy of 97.6%, indicating effective detection of TEC depletions.
- No Assumptions Required: Unlike traditional mathematical algorithms, ML techniques do not require assumptions or threshold values, making them more flexible.

Disadvantages

- Data Dependency: The performance of ML techniques heavily relies on the quality and quantity of training data.
- Complexity: Implementing machine learning models can be complex and may require significant computational resources.

2.2 MACHINE LEARNING APPROACH FOR DETECTION OF PLASMA DEPLETIONS FROM TEC

YEAR:**2024**

AUTHORS: Chandan Kapil, Gopi K. Seemala

Abstract

The paper addresses the issue of ionospheric delay affecting trans-ionospheric radio communication, particularly for navigation systems that rely on satellite signals. It notes that most ionospheric delays are estimated using dual frequency GNSS receivers, except during irregularities and equatorial plasma bubbles. Plasma bubbles are identified as decreases in total electron content (TEC) due to large-scale irregularities in the equatorial ionosphere. The study explores various machine learning (ML) techniques for detecting TEC depletions, with the Random Forest Regressor demonstrating superior performance. The RFM achieved a training accuracy of 97.6% and a positive predictive value (PPV) of 96.8%. The results suggest that ML techniques can outperform traditional mathematical algorithms in detecting TEC depletions, with potential for further advancements.

Advantages

- High Accuracy: The Random Forest Regressor achieved a training accuracy of 97.6%, indicating effective detection of TEC depletions.
- No Assumptions Required: ML techniques do not require assumptions or threshold values, making them more flexible than traditional methods.

Disadvantages

- Data Dependency: The performance of ML techniques heavily relies on the quality and quantity of training data.
- Complexity: Implementing machine learning models can be complex and may require significant computational resources.

2.3 MODELLING AND FORECASTING OF IONOSPHERIC TEC IRREGULARITIES OVER A LOW LATITUDE GNSS STATION

YEAR: 2020

AUTHORS: G. Sivavaraprasad, D. Venkata Ratnam, M. Sridhar,

K. Sivakrishna

Abstract

The study aims to leverage a systematic combination of statistical procedures, namely Principal Component Analysis (PCA) and Linear Models, to develop a local climatological model for ionospheric Total Electron Content (TEC) irregularities. Additionally, Auto Regressive Moving Average (ARMA) and Neural Network (NN) models are employed to forecast ionospheric irregularities over the Bengaluru region, India. The dataset from the Bengaluru International GNSS Service (IGS) station, covering an 8-year period, is utilized to implement the proposed algorithm. The study retrieves the main components of the time series using PCA, identifies and analyzes trends and cycles in the PCA residuals to model un-modeled irregularities using Linear models, and fits the residual series using ARMA/NN models for forecasting. The ionospheric TEC irregularities, measured in TECU/hour, are investigated during the ascending and descending phases of the 24th solar cycle. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values are used to validate the proposed models in forecasting ionospheric TEC irregularities during both geomagnetic quiet and storm days (12–14 October 2016). The NN model shows an MAE of 0.5 TECU/hour during geomagnetic quiet periods and 0.98 TECU/hour during

disturbed periods. The RMSE values are 0.73 TECU/hour for the ARMA model and 0.67 TECU/hour for the NN model, indicating that the NN model performs better in forecasting during geomagnetic quiet periods. The proposed model can serve as a climatological tool for ionospheric irregularities in low latitude regions.

Advantages

• High Accuracy: The Neural Network model demonstrated lower MAE and RMSE values, indicating better forecasting performance, especially during geomagnetic quiet periods.

• Comprehensive Approach: The study combines multiple statistical methods (PCA, Linear Models, ARMA, and NN) to create a robust model for forecasting TEC irregularities.

Disadvantages

• Complexity of Implementation: The combination of multiple statistical methods may complicate the implementation and understanding of the model.

• Data Dependency: The accuracy of the model is heavily reliant on the quality and comprehensiveness of the input data.

2.4 PERFORMANCE ANALYSIS OF NEURAL NETWORKS WITH IRI-2016 AND IRI-2012 MODELS OVER INDIAN LOW-LATITUDE GPS STATIONS

YEAR: 2020

AUTHORS: Lakshmi Mallika I, D. Venkata Ratnam, Saravana Raman, G. Sivavaraprasad

Abstract

The Global Positioning System (GPS) applications are highly vulnerable to the ionospheric space weather effects. Modelling and forecasting the ionospheric effects such as time delays for GPS signals are important for real-time alerts of space weather effects on GPS services. In the present work, the performance of the Neural Networks (NN) model is compared with International Reference Ionosphere (IRI) models. The ionospheric Total Electron Content (TEC) observations have been collected during the year 2015, during the descending phase of the 24th solar cycle, over three Indian low latitude GPS stations namely, Bengaluru (Geographic coordinates: 13.02°N and 77.57°E), near the geomagnetic equator, Guntur (Geographic coordinates: 16.37°N and 80.37°E), which is at the Equatorial Ionization Anomaly (EIA), and Lucknow (Geographic coordinates: 26.83°N and 80.92°E), which is beyond the EIA region. The performance of the NN model in predicting the ionospheric TEC values is compared with IRI (IRI-2012 and IRI-2016) models during the test period, October–December 2015, over three Indian low latitude regions. It is observed that IRI models (IRI-2012 and IRI-2016) have shown more temporal differences with GPS-VTEC during sunrise hours compared to sunset hours over the three low latitude regions. The performance of the IRI-2016 model has been apparently better than the IRI-2012 model. However, it is observed that the IRI-2016 model has large discrepancies over Bengaluru and Guntur stations due to high VTEC fluctuations at equatorial and low latitudes. The NN models have well predicted

the measured diurnal mean VTEC variations with less errors, ±5 TECU, while the differences of IRI models are ±15 TECU over all three stations. Later, GPS data for 10 years, 2009–2018, is collected over the Bengaluru station during the 24th solar cycle. The performance of the NN model is validated during 2016, 2017, and 2018 years over the Bengaluru GPS station. The error measurements and experimental results reported that the measured GPS-VTEC values are well predicted by the NN model compared to the IRI-2016 model over equatorial and low latitude GPS stations.

Advantages

• Improved Prediction: The Neural Network model demonstrated better prediction accuracy for ionospheric TEC values compared to IRI models, especially during diurnal variations.

• Lower Error Rates: The NN model achieved lower error rates (±5 TECU) compared to IRI models (±15 TECU), indicating its effectiveness in forecasting.

Disadvantages

• Model Discrepancies: The IRI-2016 model showed large discrepancies at certain stations (Bengaluru and Guntur) due to high VTEC fluctuations, indicating limitations in its applicability.

• Complexity of Neural Networks: Implementing and training Neural Network models can be complex and may require significant computational resources.

2.5 PERFORMANCE EVALUATION OF NEURAL NETWORK TEC FORECASTING MODELS OVER EQUATORIAL LOW-LATITUDE INDIAN GNSS STATION

YEAR: 2020

AUTHORS: G. Sivavaraprasad, V.S. Deepika, D. Sreenivasarao,

M. Ravi Kumar, M. Sridhar

Abstract

Global Positioning System (GPS) services could be improved through the prediction of ionospheric delays for satellite-based radio signals. With respect to latitude, longitude, local time, season, solar cycle, and geomagnetic activity, the Total Electron Content (TEC) exhibits significant variations in both time and space. These temporal and spatial TEC variations, driven by interplanetary space weather conditions such as solar and geomagnetic activities, can degrade the communication and navigation links of GPS. Hence, in this paper, the performance of TEC forecasting models based on Neural Networks (NN) has been evaluated to forecast (1-h ahead) ionospheric TEC over the equatorial low latitude Bengaluru Global Navigation Satellite System (GNSS) station, India. The VTEC data is collected for 2009–2016 (8 years) during the current 24th solar cycle. The input space for the NN models comprises the solar Extreme UV flux, F10.7 proxy, geomagnetic planetary A index (AP) index, sunspot number (SSN), disturbance storm time (DST) index, solar wind speed, solar wind proton density, and Interplanetary Magnetic Field (IMF). The performance of NN-based TEC forecast models and the International Reference Ionosphere, IRI-2016 global TEC model has been evaluated during the testing period in 2016. The NN-based model driven by all the inputs, which is a NN unified model (NNunq), has shown better accuracy with a Mean Absolute Error (MAE) of 3.15 TECU, Mean Square Deviation (MSD) of 16.8, and Mean Absolute Percentage Error (MAPE) of 19.8%, and is 1–25% more accurate than the other NN-based TEC forecast models (NN1, NN2, and NN3) and the IRI-2016 model. The NNunq model has a

lower Root Mean Square Error (RMSE) value of 3.8 TECU and the highest goodness-of-fit (R^2) with 0.85. The experimental results imply that the NNunq/NN1 model forecasts ionospheric TEC accurately across the equatorial low-latitude GNSS station, and the performance of the IRI-2016 model is necessarily improved as its forecast accuracy is limited to 69–70%.

Advantages

• Improved Prediction: The NNunq model demonstrated better prediction accuracy for ionospheric TEC values compared to IRI models, especially during diurnal variations.

• Lower Error Rates: The NNunq model achieved lower error rates (MAE of 3.15 TECU) compared to IRI models, indicating its effectiveness in forecasting.

Disadvantages

• Model Discrepancies: The IRI-2016 model showed large discrepancies at certain stations due to high VTEC fluctuations, indicating limitations in its applicability.

• Complexity of Neural Networks: Implementing and training Neural Network models can be complex and may require significant computational resources.

2.6 GLOBAL IONOSPHERIC TOTAL ELECTRON CONTENT SHORT-TERM FORECAST BASED ON LIGHT GRADIENT BOOSTING MACHINE, EXTREME GRADIENT BOOSTING, AND GRADIENT BOOST REGRESSION

YEAR: 2024

AUTHORS: Suneetha Emmela, V. Rama Lahari, B. Anusha, D. Bhavana, Yury V. Yasyukevich, Vladislav V. Demyanov, D. Venkata Ratnam

Abstract

Total Electron Content (TEC) forecasting using machine learning has been extensively preferred in characterizing the spatio-temporal variability of the ionosphere, to support space communication and navigation applications. Although a variety of machine learning methods have evolved, there is still a greater persistence in the prediction of ionospheric TEC due to adverse space weather impacts and the complex behavior of the ionosphere. Hence, more sophisticated short-term ionospheric TEC forecasting models should be developed for better prediction accuracy. The present study investigated and evaluated the performance of three models – Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost Regressor), and Gradient Boost Regression (GBR) algorithms using the dataset for 25 years (1997–2021) GNSS Earth Observation Network System (GEONET) Global mean TEC time-series data. For better prediction accuracy, the data of geomagnetic storm indices Dst and Mg-II index were also utilized in training the three models. One year of data for both high (2015) and low (2020) solar activity were used to test the three models. The prediction plots are used to visualize the level of prediction error for the three algorithms. The statistical results illustrate that the LightGBM model performed better in predicting TEC with an RMSE value of 2.25 TECU (2015) and 0.52 TECU (2020) for high and low solar activity years respectively for global mean TEC data and with an RMSE of 2.34 TECU at Japan grid point location of (34.95°N and 134.05°E). In turn, the XGBoost Regressor and GBR

models provide an absolute TEC forecasting with an RMSE of 2.76 and 2.59 TECU (2015) and 0.60 and 0.55 TECU (2020) correspondingly for global data and as 2.39 and 2.93 TECU for grid point location.

Advantages

• Improved Prediction: The LightGBM model demonstrated better prediction accuracy for ionospheric TEC values compared to XGBoost Regressor and GBR models, especially during high solar activity years.

• Lower Error Rates: The LightGBM model achieved lower error rates (RMSE of 2.25 TECU) compared to XGBoost Regressor and GBR models, indicating its effectiveness in forecasting.

Disadvantages

• Model Discrepancies: The XGBoost Regressor and GBR models showed larger discrepancies in predicting TEC values, indicating limitations in their applicability.

• Complexity of Machine Learning Models: Implementing and training machine learning models can be complex and may require significant computational resources.

2.7 BI-LSTM BASED VERTICAL TOTAL ELECTRON CONTENT PREDICTION AT LOW-LATITUDE EQUATORIAL IONIZATION ANOMALY REGION OF SOUTH INDIA

YEAR: 2024

AUTHORS: Veera Kumar Maheswaran, James A. Baskaradas, Raju Nagarajan, Rajesh Anbazhagan, Sriram Subramanian, Venkata Ratnam, Rupesh M. Das

Abstract

In this study, we consider a Bi-directional Long Short Term Memory (Bi-LSTM) model based Vertical Total Electron Content (VTEC) prediction over Thanjavur (Geographic 10.72˚ N, 79.01˚ E, Geomagnetic 2.34˚ N, 152.19˚ E) Global Positioning System (GPS) station. This station is located at low latitude Equatorial Ionization Anomaly (EIA) region of 2˚ geomagnetic dip latitude and has unique ionospheric dynamics. In this region, the VTEC prediction is crucial and challenging for space weather and the Sixth Generation (6G) Internet of Space (IoS) application to support early warning systems and future spatial data transmissions. A Deep Learning (DL) model based on Bi-LSTM was developed and trained for F10.7 and Dst index for predicting the VTEC. This study highlights the prediction of VTEC for any day that includes solstice and equinox time frames. The Bi-LSTM has an improvement of 28% in mean absolute error (MAE), 48% in mean square error (MSE), and 24% in root mean square error (RMSE) as compared to the conventional Long Short Term Memory (LSTM) network. Hence, this Bi-LSTM model can be helpful to predict the VTEC in the EIA region and may be helpful to extrapolate over the unmeasured grid region of ocean and land.

Advantages

• Enhanced Accuracy: The Bi-LSTM model showed significant improvements in prediction accuracy compared to traditional LSTM models, with reductions in MAE, MSE, and RMSE.

• Effective for Unique Dynamics: The model is specifically designed to handle the unique ionospheric dynamics of the EIA region, making it suitable for local applications.

Disadvantages

• Complexity of Implementation: Bi-LSTM models can be complex to implement and require significant computational resources for training and validation.

• Data Dependency: The model's performance is heavily reliant on the quality and quantity of input data, which may not always be available, especially during extreme space weather events.

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 HARDWARE AND SOFTWARE REQUIREMENTS

To efficiently develop and deploy machine learning models for Total Electron Content prediction, a well-equipped computing environment is essential. The hardware and software specifications play a crucial role in handling large datasets, running machine learning algorithms, and ensuring smooth performance.

3.1.1 Hardware Requirements

A robust computing system is necessary for efficient model training and real-time predictions. The recommended hardware includes:

- Processor: A multi-core processor such as Intel Core i5/i7 (10th Gen or later) is required. These processors provide the necessary computational power to handle complex calculations and data processing.
- RAM: At least 8GB of RAM is required to run standard machine learning models. However, for handling large datasets and deep learning applications, 16GB or more is recommended to ensure seamless processing and avoid memory-related issues.
- Storage: A Solid-State Drive (SSD) with at least 256GB is preferred for faster data access and storage, particularly when dealing with large datasets.

3.1.2. Software Requirements

The development environment should be compatible with machine learning tools and frameworks. The key software components include:

- Operating System: The models can be developed on Windows 10/11, Ubuntu 20.04+, or macOS Monterey. Linux-based systems like Ubuntu are preferred for better compatibility with open-source machine learning tools.
- Python Version: The implementation requires Python 3.8 or later, as it supports the latest machine learning libraries and frameworks.
- IDE (Integrated Development Environment): For efficient coding and experimentation, Google Colab Notebook is recommended. It provides cloud-based computing with GPU support, enabling faster training of deep learning models without requiring a high-end local machine.

3.1.3. Required Libraries

Various Python libraries are essential for data handling, visualization, and machine learning model development. The key libraries include:

- Data Handling:
    - Pandas: Used for data manipulation and preprocessing. It provides efficient data structures like DataFrames to organize and analyze tabular data.
    - NumPy: Essential for numerical computations, including matrix operations and statistical analysis.
- Visualization:
    - Matplotlib: A fundamental library for creating static, animated, and interactive visualizations. It helps in plotting trends and patterns in ionospheric data.
    - Seaborn: Built on Matplotlib, this library provides advanced visualization techniques, making it easier to analyze correlations and distributions.

- Machine Learning:
    - Scikit-learn: A versatile machine learning library that includes various classification, regression, and clustering algorithms.
    - XGBoost Regressor: An optimized gradient boosting framework widely used for high-performance predictive modeling. It enhances model accuracy while reducing overfitting.
    - Random Forest Regressor: A powerful ensemble learning algorithm used for regression tasks. It combines multiple decision trees to improve prediction accuracy and robustness.

By ensuring these hardware and software requirements are met, researchers and developers can effectively build and optimize EPB prediction models, contributing to improved navigation and communication reliability.

CHAPTER 4

MODULE DESCRIPTION

4.1 DATA PREPROCESSING:

- Gather the dataset: IONOLAB, OMNIWEB and IRI 2020
- Read the dataset containing columns: DATE, TIME, TRUE TEC, KP, SSN, AP, F10.7, IRI_TEC VALUES.
- Drop irrelevant columns: DATE and IRI_TEC and convert date into datetime
- Handle Missing Values: Fill missing values using interpolation or mean/mode imputation.

4.2 FEATURE ENGINEERING

- Key parameters influencing TEC variations (e.g., solar flux (F10.7), Kp, Ap, SSN) are selected.
- Time-based features (hourly, daily, seasonal trends) are added to improve model learning.
- Generate Polynomial features, Use Polynomial Features to capture nonlinear relationships.

4.3 MODEL TRAINING

- Split data into training and testing (train the data before the date of prediction)
- Select historical TEC data to train models.
- Extract input features (KP, SSN, AP, F10.7) and target variable (TRUE TEC).
    - **KP Index (Kp)**: A global geomagnetic activity index that indicates the level of disturbances in the Earth's magnetic field due to solar wind.

- **Sunspot Number (SSN)**: A measure of solar activity based on the number of sunspots, which correlates with increased solar radiation and magnetic activity.
- **Planetary Ap Index (Ap)**: An index that quantifies geomagnetic activity based on magnetic field measurements from multiple locations worldwide.
- **F10.7 Solar Flux**: A measure of solar radio emissions at 10.7 cm wavelength, indicating solar activity and its potential impact on the ionosphere.

- Train and optimize Random Forest Regressor and XGBoost Regressor models to improve prediction accuracy.

4.4 PREDICTION

1. Predicting TEC Values

- Trained machine learning models are used to forecast TEC levels at future timestamps based on historical data and space weather parameters.
- Features such as solar activity indices (e.g., solar flux, sunspot number), geomagnetic conditions, and ionospheric parameters are used for predictions.

2. Predicting TEC During Extreme Events

TEC variations can be significantly influenced by natural and space weather events such as earthquakes, solar flares, and solar storms. Predictive models can be adapted to forecast TEC changes during these specific events:

- **Earthquakes:** Sudden ionospheric disturbances have been observed before and after major seismic events. Predicting these variations can assist in early warning systems.

- **Solar Flares:** Intense bursts of solar radiation impact the ionosphere, causing abrupt changes in TEC. Machine learning models help estimate these disturbances.
- **Solar Storms:** Geomagnetic storms trigger ionospheric anomalies, which can be forecasted to minimize technological disruptions.

4.5 MODEL EVALUATION AND VISUALISATION

- Compute Performance Metrics
    - Mean Squared Logarithmic Error (MSLE): Measures the squared logarithmic differences between actual and predicted values, penalizing underestimations more.
    - R-squared Score (R^2): Indicates how well the model explains variance in the target variable, ranging from 0 (poor) to 1 (perfect).
    - Symmetric mean absolute percentage error (SMAPE): Calculates the percentage error between actual and predicted values, treating over- and under-predictions equally.
- Visualize Predictions vs. Actual True TEC values
- Use Matplotlib & Seaborn to plot TEC variations over time.
- Compare predicted TEC and actual TEC with IRI2020.

4.6 TEC PREDICTION USING SOLAR ACTIVITY AND MACHINE LEARNING

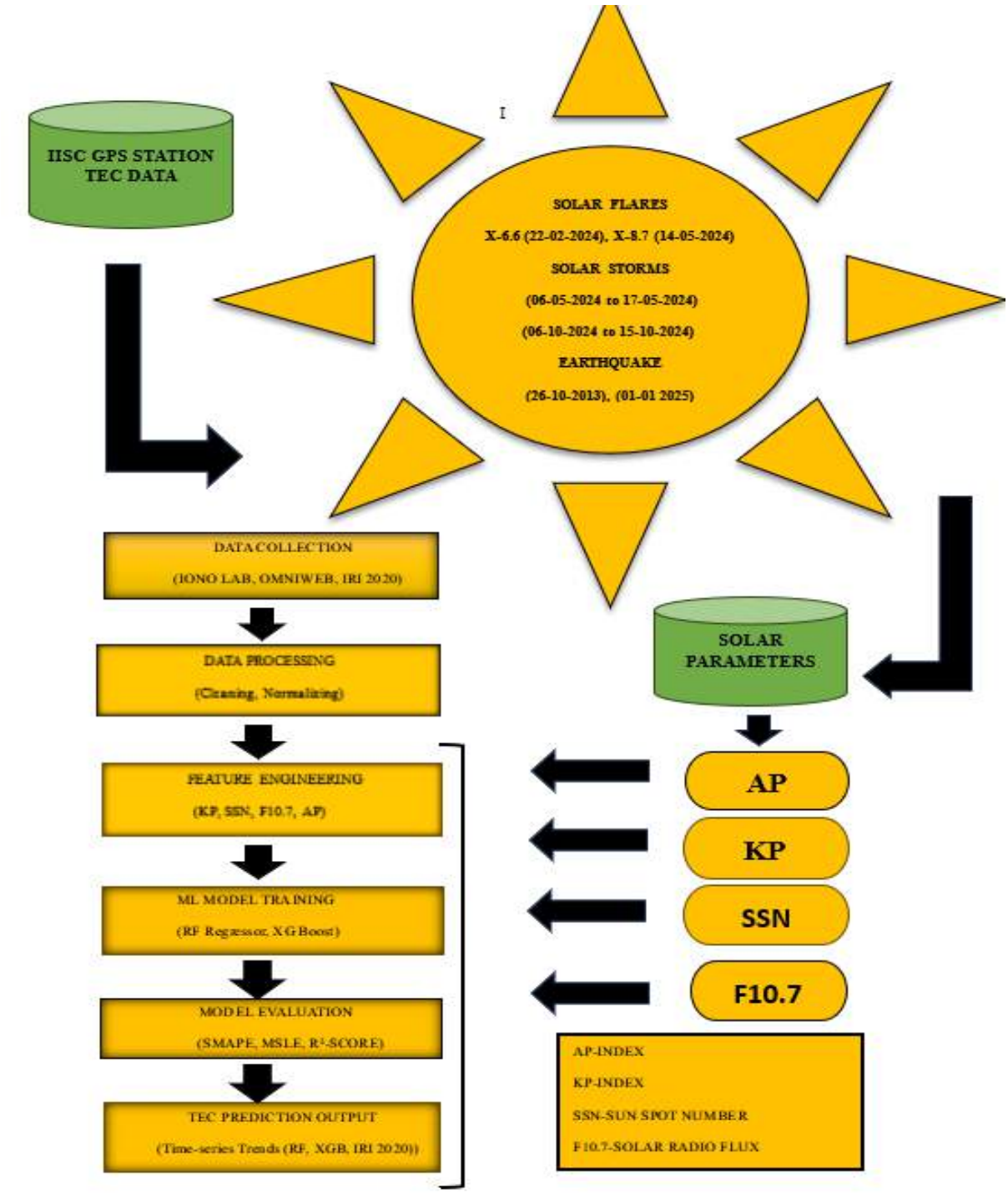

Figure 4.1: Workflow for TEC Prediction Using Machine Learning Models

4.6.1 Data Sources

The prediction framework is driven by data collected from both terrestrial and solar observations:

- **TEC Data** is obtained from **IISC GPS Station**, representing ionospheric electron density variations over time.
- **Solar Parameters** are acquired from reputable sources like **IONO Lab**, **OMNIWeb**, and **IRI 2020** models. These include:
    - **AP Index** – Reflects daily geomagnetic activity levels.
    - **KP Index** – Measures short-term geomagnetic fluctuations.
    - **SSN (Sunspot Number)** – Indicates solar activity based on visible sunspots.
    - **F10.7 Index** – A proxy for solar EUV emissions, affecting ionization levels in the ionosphere.

4.6.2 Solar Activity Events

To capture external influences on TEC variation, the model integrates significant **solar-terrestrial events**:

- **Solar Flares:**
  Explosive bursts of radiation from the sun, with notable events:
    - X6.6 on 22-Feb-2024
    - X8.7 on 14-May-2024
- **Solar Storms:**
  Periods of elevated geomagnetic activity that can disturb the ionosphere:
    - 06-May-2024 to 17-May-2024
    - 06-Oct-2024 to 15-Oct-2024

- **Earthquakes:**
  Geophysical disturbances which may have secondary ionospheric effects:
  - 26-Oct-2013
  - 01-Jan-2025

These events help analyze correlations between solar/terrestrial disturbances and ionospheric TEC variations.

4.6.3 Workflow Pipeline

The end-to-end pipeline for TEC prediction includes:

- **Data Collection:**
  Acquiring TEC and solar parameter data from global scientific repositories and real-time monitoring systems.
- **Data Processing:**
  Involves cleaning noisy data, handling missing values, and normalizing variables to prepare for modeling.
- **Feature Engineering:**
  Constructing meaningful inputs from raw solar indices (KP, SSN, F10.7, AP) to enhance model learning.
- **Model Training:**
  Applying supervised machine learning algorithms:
  - **Random Forest Regressor (RF)**
  - **XGBoost Regressor**
    These models capture nonlinear patterns and interactions among features.

- **Model Evaluation:**

  Assessing accuracy using:

  - **SMAPE** (Symmetric Mean Absolute Percentage Error)
  - **MSLE** (Mean Squared Logarithmic Error)
  - **R^2 Score** (Coefficient of Determination)

- **Prediction Output:**

  Produces forecasted TEC trends compared against the IRI 2020 model, highlighting deviations and model strengths.

CHAPTER 5

IMPLEMENTATION

5.1 MACHINE LEARNING MODELS

The study employed two powerful machine learning algorithms — Random Forest Regressor and XGBoost Regressor — to predict TEC fluctuations. Both models are known for handling complex, nonlinear data, making them suitable for capturing the dynamic nature of ionospheric variations during solar and geomagnetic disturbances.

5.1.1 Random Forest Regressor Algorithm

Introduction

Random Forest Regressor is a robust ensemble learning algorithm widely used for classification and regression tasks. It is particularly effective in handling complex, nonlinear data, making it suitable for applications such as ionospheric modeling, weather prediction, and financial forecasting. By aggregating multiple decision trees, Random Forest Regressor enhances predictive accuracy while reducing the risk of overfitting.

Working Mechanism

Random Forest Regressor operates based on the principles of **Bootstrap Aggregation (Bagging)** and **Feature Randomness**, which contribute to its efficiency and robustness.

1. **Bootstrap Sampling**: The algorithm creates multiple subsets of the original dataset by randomly sampling with replacement. Each subset is used to train an individual decision tree.

2. **Feature Subsetting**: At each node split, a random subset of features is considered instead of the entire feature set, which reduces correlation between trees and improves generalization.

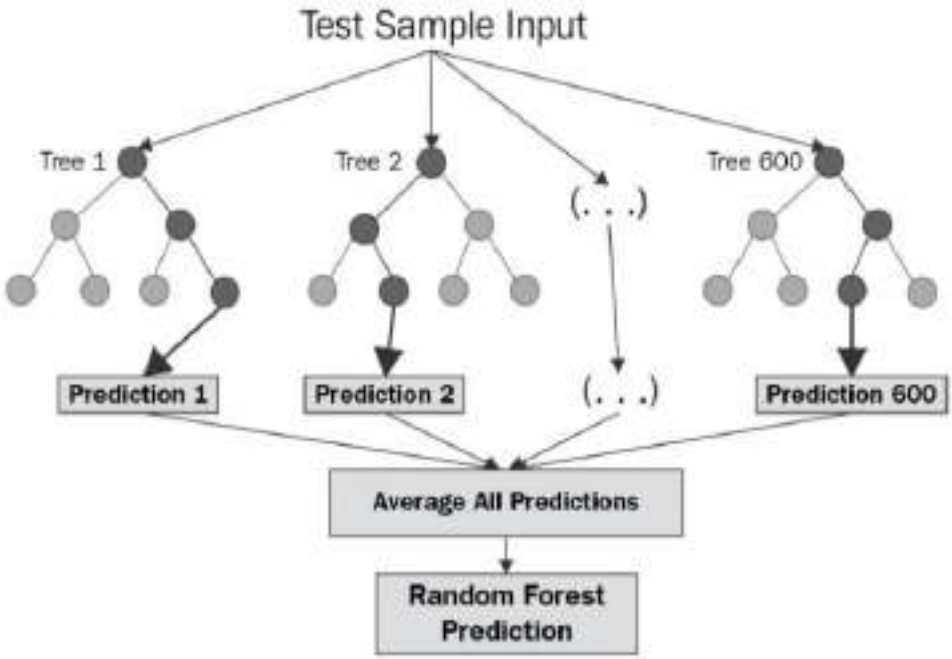

Figure 5.1: Random Forest Model => Prediction from Multiple Decision Trees

3. **Tree Independence**: Each decision tree is trained independently, making the model resistant to noise and overfitting.
4. **Prediction Aggregation**:
    - **For Regression**: The final prediction is obtained by averaging the outputs of all trees.

Advantages of Random Forest Regressor

- **Handles Nonlinearity and Noisy Data**: Suitable for capturing complex relationships in datasets.
- **Robust to Overfitting**: The averaging mechanism ensures that individual tree biases do not dominate the final prediction.

Key Hyperparameters and Their Roles

To optimize model performance, various hyperparameters can be tuned:

- **n_estimators**: Defines the number of trees in the forest. A larger number improves stability but increases computational cost.
- **max_depth**: Limits the depth of each tree to prevent overfitting.
- **min_samples_split**: Minimum number of samples required to split a node, controlling tree complexity.

Limitations of Random Forest Regressor

Despite its advantages, Random Forest Regressor has some limitations:

- **Computational Complexity**: Training a large number of trees increases processing time and memory usage.
- **Lower Interpretability**: Unlike individual decision trees, the ensemble nature makes it harder to interpret results.

Random Forest Regressor is a powerful machine learning algorithm that excels in handling complex datasets with nonlinear relationships. Its ensemble approach ensures robustness, making it a preferred choice for tasks requiring high accuracy and generalization. By tuning hyperparameters appropriately, it can be optimized for various real-world applications, including the prediction of ionospheric variations in space weather studies.

5.1.2 XGBoost Regressor Algorithm

Introduction

XGBoost Regressor (Extreme Gradient Boosting) is an efficient and high-performance gradient boosting algorithm known for its speed and accuracy. Unlike Random Forest Regressor's Regressor approach, XGBoost Regressor follows a boosting strategy, where trees are built sequentially, with each new tree

learning from the errors made by the previous ones. This iterative approach enhances model accuracy, making it highly effective for structured datasets and predictive analytics.

Working Mechanism

XGBoost Regressor improves traditional gradient boosting methods by incorporating several optimization techniques:

- **Boosting Strategy**: Trees are trained sequentially, correcting the errors of the previous models to refine predictions.
- **Gradient-Based Learning**: Errors from prior models are analyzed using gradient descent principles to improve performance in subsequent iterations.

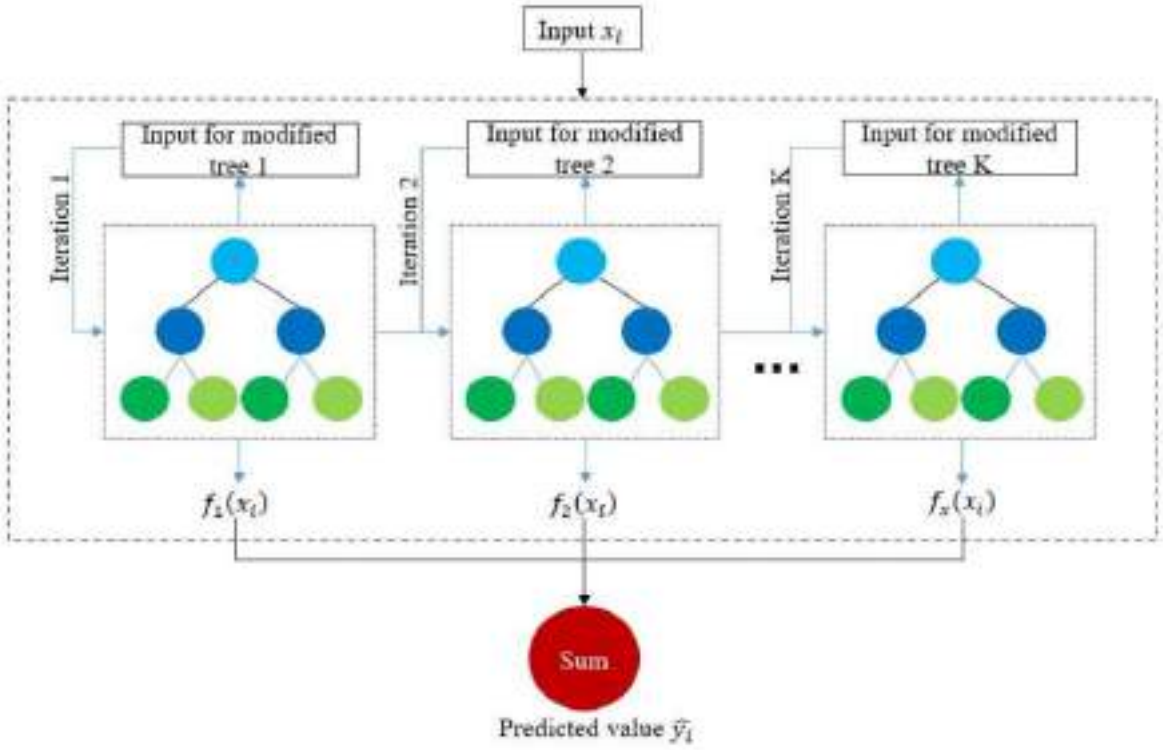

Figure 5.2: Schematic Representation of XGBoost Model

- **Regularization (L1 & L2)**: XGBoost Regressor integrates both Lasso (L1) and Ridge (L2) regularization terms to penalize overly complex models, reducing overfitting.
- **Handling Missing Data**: It automatically determines the optimal direction for missing values at each split, making it resilient to incomplete datasets.

Advantages of XGBoost Regressor

- **High Accuracy**: Due to its sequential boosting approach, it often outperforms other ensemble methods.
- **Efficient Computation**: Optimized for speed with parallel processing and tree pruning.

Key Hyperparameters and Their Roles

To optimize model performance, the following hyperparameters were tuned:

- **n_estimators**: Number of boosting rounds; higher values improve accuracy but increase training time.
- **learning_rate**: Controls the step size at each iteration. Lower values ensure gradual learning and better generalization.
- **max_depth**: Limits the depth of trees to balance complexity and overfitting.

Limitations of XGBoost Regressor

- **Computationally Expensive**: Training can be slower compared to simpler models, especially on large datasets.
- **Memory Intensive**: Requires significant RAM for large-scale datasets due to multiple tree-building iterations.

- **Prone to Overfitting**: If not carefully regularized, XGBoost Regressor can overfit the training data, especially with a high number of trees.

Both **Random Forest Regressor** and **XGBoost Regressor** serve as powerful machine learning algorithms for TEC prediction. While Random Forest Regressor provides robust performance through ensemble learning, XGBoost Regressor enhances accuracy through sequential gradient boosting.

5.2 EVALUATION METRICS FOR PREDICTING IONOSPHERIC TEC

(During X-Class Solar Flares, Solar storms and Earthquakes Using Machine Learning Models)

5.2.1 Symmetric Mean Absolute Percentage Error (SMAPE)

SMAPE is a statistical measure used to assess the accuracy of regression models by comparing predicted values with actual values in a symmetric manner. Unlike traditional Mean Absolute Percentage Error (MAPE), which can be biased by extreme values, SMAPE ensures that both overestimation and underestimation are treated equally.

Formula:

$$\textbf{SMAPE} = \frac{1}{n} * \Sigma \left[\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \right] * 100$$

Interpretation:

- Lower SMAPE values indicate better model performance.
- Higher SMAPE values suggest greater discrepancies between predicted and actual values.

5.2.2 Coefficient of Determination (R^2)

The Coefficient of Determination (R^2) measures how well the predicted values explain the variability of the actual values. It is commonly used to evaluate

regression models and represents the proportion of variance in the dependent variable that is predictable from the independent variables.

Formula:

$$\mathbf{R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2]}$$

Interpretation:

- $R^2 = 1$ → Perfect fit, meaning the model explains all the variability in the data.
- R^2 close to 1 → High explanatory power of the model.
- $R^2 = 0$ → The model does not explain any variability better than a simple mean prediction.
- Negative R2R^2 → The model performs worse than a naive mean-based prediction.

R^2 is particularly useful for assessing the goodness of fit of models predicting Total Electron Content (TEC) variations during extreme solar activity, such as X-class solar flares.

5.2.3 Mean Squared Logarithmic Error (MSLE)

MSLE is an error metric that calculates the squared logarithmic differences between actual and predicted values. It is useful when dealing with skewed data, where underestimation errors are more critical than overestimation errors.

Formula:

$$\mathbf{MSLE = (1/n) * \Sigma [(\log(1 + y_i) - \log(1 + \hat{y}_i))^2]}$$

Interpretation:

- Lower MSLE values indicate better model performance.

- It penalizes under-predictions more than over-predictions, making it useful for applications where missing a peak in TEC values is more critical than overestimating.
- It is beneficial for models predicting ionospheric TEC, where values can fluctuate significantly due to solar activity.

Conclusion

These three Evaluation metrics provide a robust assessment of machine learning models like Random Forest Regressor and XGBoost Regressor when predicting ionospheric TEC variations during X-class solar flares. SMAPE ensures balanced error measurement, R^2 evaluates the model's explanatory power, and MSLE emphasizes the impact of underestimations, making them collectively effective for performance analysis.

5.3 SOLAR FLARES

5.3.1 Prediction of Tec During the X6.3 Solar Flare

On February 22, 2024, a significant X6.3-class solar flare erupted from Active Region 3590. The event commenced at 22:08 UTC, peaked at 22:34 UTC, and concluded at 22:43 UTC. In this study, we predicted the Total Electron Content (TEC) during the X6.3 solar flare, as well as four days before and four days after the event, using Random Forest Regressor and XGBoost Regressor models.

Table 5.1: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during X6.3 solar flare

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
18-02-2024	16.97	22.24	40.34	0.85	0.88	0.309	0.047	0.06	0.22
19-02-2024	15.35	24.91	34.55	0.9	0.94	0.453	0.034	0.024	0.13
20-02-2024	12.93	23.54	35.08	0.93	0.96	0.469	0.02	0.017	0.15
21-02-2024	11.53	21.79	32.48	0.97	0.97	0.496	0.015	0.013	0.13
22-02-2024	17.25	29.1	38.16	0.89	0.9	0.47	0.048	0.056	0.17
23-02-2024	9.627	23.44	30.28	0.98	0.97	0.592	0.009	0.017	0.11
24-02-2024	9.364	24.86	32.12	0.98	0.96	0.591	0.012	0.019	0.14
25-02-2024	17.49	19.8	37.9	0.85	0.89	0.367	0.068	0.057	0.18
26-02-2024	9.792	23.02	31.48	0.97	0.97	0.501	0.014	0.011	0.12
AVERAGE	**13.37**	**23.63**	**34.71**	**0.92**	**0.94**	**0.472**	**0.03**	**0.03**	**0.15**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on metrics such as Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis provided valuable insights into the models' predictive capabilities under varying solar conditions around February 22, 2024.

On February 22, the day of the X6.3 solar flare, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 17.251, MSLE of 0.047, and R^2 of 0.889, while XGBoost Regressor remained the best performer with a SMAPE of 29.100, MSLE of 0.055, and R^2 of 0.902. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 38.159, MSLE of 0.169, and a significantly lower R^2 of 0.470.

Overall, the analysis indicates that Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC values during the X6.3 solar flare event. XGBoost Regressor emerged as the best performing model across all metrics, demonstrating the lowest SMAPE and

MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the flare event on February 22, highlighting the superior adaptability of machine learning models in capturing TEC variations under extreme solar conditions.

5.3.2 Prediction of Tec During the X8.79 Solar Flare

On May 14, 2024, a significant X8.7-class solar flare erupted from Active Region 3664. The event commenced at 16:46 UTC, peaked at 16:51 UTC, and concluded at 17:02 UTC. In this study, we predicted the Total Electron Content (TEC) during the X8.7 solar flare, as well as four days before and four days after the event, using Random Forest Regressor and XGBoost Regressor models.

Table 5.2: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during X8.79 solar flare

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
10-05-2024	18.85	32.6	38.56	0.92	0.96	0.581	0.115	0.056	0.21
11-05-2024	59.92	64.75	66.95	0.72	0.71	0.539	0.951	0.832	0.97
12-05-2024	21.6	31.83	38.11	0.91	0.88	0.591	0.098	0.08	0.27
13-05-2024	11.35	20.79	29.42	0.95	0.91	0.585	0.021	0.034	0.11
14-05-2024	9.704	12.48	24.03	0.96	0.97	0.563	0.018	0.011	0.08
15-05-2024	6.94	15.68	22.1	0.97	0.96	0.582	0.015	0.017	0.07
16-05-2024	12.1	19.49	24.7	0.95	0.93	0.501	0.015	0.018	0.08
17-05-2024	17.49	27.2	31.95	0.94	0.92	0.519	0.04	0.037	0.13
18-05-2024	14.24	18.17	27.1	0.96	0.94	0.415	0.014	0.017	0.11
AVERAGE	**19.13**	**27**	**33.66**	**0.92**	**0.91**	**0.542**	**0.143**	**0.122**	**0.23**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on metrics such as Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis provided valuable insights

into the models' predictive capabilities under varying solar conditions around May 14, 2024.

On May 14, the day of the X8.7 solar flare, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 9.704, MSLE of 0.018, and R^2 of 0.957, while XGBoost Regressor remained the best performer with a SMAPE of 12.484, MSLE of 0.011, and R^2 of 0.974. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 24.028, MSLE of 0.079, and a significantly lower R^2 of 0.563. Following the flare, model performance gradually improved.

Overall, the analysis indicates that Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC values during the X8.7 solar flare event. XGBoost Regressor emerged as the best performing model across all metrics, demonstrating the lowest SMAPE and MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the flare event on May 14, highlighting the superior adaptability of machine learning models in capturing TEC variations under extreme solar conditions.

5.4 SOLAR STORMS

5.4.1 Prediction of Tec During the Solar Storm (May 6, 2024 to May 17, 2024)

On May 6, 2024, a significant solar storm event began, lasting through May 17, 2024. In this study, we predicted the Total Electron Content (TEC) during the solar storm event, as well as four days before and four days after using Random Forest Regressor and XGBoost Regressor models.

Table 5.3: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during solar storm.

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
06-05-2024	12.67	9.893	23.9	0.949	0.968	0.71	0.03	0.016	0.11
07-05-2024	19.65	20.52	31.23	0.81	0.847	0.631	0.07	0.077	0.13
08-05-2024	15.04	12.87	27.62	0.884	0.924	0.607	0.03	0.026	0.1
09-05-2024	15.16	12.19	31.72	0.937	0.958	0.606	0.03	0.023	0.13
10-05-2024	18.12	14.61	38.56	0.922	0.957	0.581	0.11	0.056	0.21
11-05-2024	58.93	59.54	66.95	0.715	0.714	0.539	0.95	0.832	0.97
12-05-2024	20.41	21.63	38.11	0.905	0.88	0.591	0.1	0.08	0.27
13-05-2024	10.98	13.86	29.42	0.949	0.912	0.585	0.02	0.034	0.11
14-05-2024	11.2	8.534	24.03	0.958	0.974	0.563	0.02	0.011	0.08
15-05-2024	9.5	9.991	22.1	0.974	0.962	0.582	0.02	0.017	0.07
16-05-2024	10.68	12.17	24.7	0.953	0.929	0.501	0.01	0.018	0.08
17-05-2024	16.91	17.33	31.95	0.945	0.92	0.519	0.04	0.037	0.13
AVERAGE	**18.27**	**17.76**	**32.52**	**0.908**	**0.912**	**0.585**	**0.12**	**0.102**	**0.2**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on metrics such as Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis provided valuable insights into the models' predictive capabilities under varying solar conditions around the solar storm event.

The performance comparison between the models showed that on May 6, 2024, the Random Forest Regressor achieved a SMAPE of 12.672, MSLE of 0.026, and an R^2 value of 0.949, while XGBoost Regressor performed slightly better with a SMAPE of 9.892, MSLE of 0.016, and an R^2 of 0.968. The IRI-2020 model, however, had a higher SMAPE of 23.900, MSLE of 0.110, and a lower R^2 of 0.709.

Similar trends were observed on May 7, where Random Forest Regressor recorded a SMAPE of 19.651, MSLE of 0.070, and R^2 of 0.810, whereas XGBoost Regressor maintained a slight edge with a SMAPE of 20.517, MSLE

of 0.077, and R^2 of 0.847. The IRI-2020 model continued to lag with a SMAPE of 31.232, MSLE of 0.127, and R^2 of 0.630. As the solar storm approached, the models' performance fluctuated slightly.

On May 8, Random Forest Regressor and XGBoost Regressor recorded SMAPE values of 15.039 and 12.871, respectively, while IRI-2020 had a much higher error at 27.617. Their MSLE values stood at 0.032 for Random Forest Regressor, 0.025 for XGBoost Regressor, and 0.100 for IRI-2020. The R^2 values followed a similar trend, with Random Forest Regressor achieving 0.884, XGBoost Regressor at 0.924, and IRI-2020 falling behind at 0.606.

On May 9, the day before the solar storm intensified, Random Forest Regressor and XGBoost Regressor recorded SMAPE values of 15.163 and 12.185, respectively, while IRI-2020 had a much higher error at 31.717. Their MSLE values stood at 0.032 for Random Forest Regressor, 0.023 for XGBoost Regressor, and 0.134 for IRI-2020. The R^2 values followed a similar trend, with Random Forest Regressor achieving 0.937, XGBoost Regressor at 0.957, and IRI-2020 falling behind at 0.606.

On May 10, as the solar storm peaked, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 18.117, MSLE of 0.114, and R^2 of 0.921, while XGBoost Regressor remained the best performer with a SMAPE of 14.614, MSLE of 0.055, and R^2 of 0.956. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 338.557, MSLE of 0.208, and a significantly lower R^2 of 0.581. Following the peak, model performance gradually improved.

On May 11, Random Forest Regressor recorded a SMAPE of 58.927, MSLE of 0.950, and R^2 of 0.715, while XGBoost Regressor maintained slightly better

accuracy with a SMAPE of 59.542, MSLE of 0.832, and R^2 of 0.714. The IRI-2020 model, however, had a SMAPE of 66.946, MSLE of 0.968, and R^2 of 0.539.

By May 12, the performance of all models stabilized, with Random Forest Regressor achieving a SMAPE of 20.408, MSLE of 0.097, and R^2 of 0.095, while XGBoost Regressor continued to excel with a SMAPE of 21.629, MSLE of 0.079, and R^2 of 0.879. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 38.110, MSLE of 0.267, and a lower R^2 of 0.591.

By May 13, the models continued to show stable performance. Random Forest Regressor achieved a SMAPE of 10.981, MSLE of 0.021, and R^2 of 0.948, while XGBoost Regressor remained ahead with a SMAPE of 13.864, MSLE of 0.033, and R^2 of 0.912. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 29.424, MSLE of 0.112, and a lower R^2 of 0.585.

By May 14, Random Forest Regressor achieved a SMAPE of 11.201, MSLE of 0.018, and R^2 of 0.957, while XGBoost Regressor continued to excel with a SMAPE of 8.533, MSLE of 0.011, and R^2 of 0.974. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 24.028, MSLE of 0.079, and a lower R^2 of 0.563.

By May 15, Random Forest Regressor achieved a SMAPE of 9.500, MSLE of 0.015, and R^2 of 0.973, while XGBoost Regressor continued to excel with a SMAPE of 9.991, MSLE of 0.017, and R^2 of 0.964. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 22.104, MSLE of 0.068, and a lower R^2 of 0.582.

By May 16, Random Forest Regressor achieved a SMAPE of 10.682, MSLE of 0.014, and R^2 of 0.952, while XGBoost Regressor continued to excel with a SMAPE of 12.165, MSLE of 0.017, and R^2 of 0.928. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 24.696, MSLE of 0.084, and lower R^2 of 0.501

By May 17, Random Forest Regressor achieved a SMAPE of 16.913, MSLE of 0.040, and R^2 of 0.944, while XGBoost Regressor continued to excel with a SMAPE of 17.333, MSLE of 0.037, and R^2 of 0.920. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 31.954, MSLE of 0.132, and a lower R^2 of 0.518.

Overall, the analysis indicates that Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC values during the solar storm event. XGBoost Regressor emerged as the best-performing model across all metrics, demonstrating the lowest SMAPE and MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the storm event, highlighting the superior adaptability of machine learning models in capturing TEC variations under extreme solar conditions.

5.4.2 Prediction of Tec During the Solar Storm (Oct 6 ,2024 to Oct 15 ,2024)

On October 10, 2024, a significant solar storm impacted Earth's geomagnetic environment, originating from a powerful coronal mass ejection (CME) associated with an M9.8-class solar flare from Active Region 3675. The storm began at 08:22 UTC, reached peak intensity at 14:45 UTC, and gradually subsided by 22:30 UTC. In this study, we predicted the Total Electron Content (TEC) during this solar storm, as well as four days before and four days after the event, using Random Forest Regressor and XGBoost Regressor models.

Table 5.4: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during solar storm.

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
06-10-2024	16.45	14.48	42.02	0.837	0.895	0.202	0.05	0.033	0.21
07-10-2024	16.2	15.13	37.7	0.869	0.907	0.33	0.04	0.053	0.18
08-10-2024	25.65	27.04	33.53	0.825	0.85	0.356	0.1	0.113	0.14
09-10-2024	33.61	30.13	50.45	0.854	0.91	0.312	0.21	0.208	0.31
10-10-2024	25.78	23.35	49.01	0.917	0.911	0.312	0.13	0.092	0.3
11-10-2024	22.85	24.58	44.31	0.905	0.866	0.455	0.12	0.121	0.3
12-10-2024	20.63	17.4	44.51	0.952	0.955	0.399	0.16	0.11	0.39
13-10-2024	18.22	16.36	44.74	0.968	0.971	0.352	0.12	0.114	0.39
14-10-2024	13.41	16.29	42.01	0.965	0.961	0.384	0.03	0.045	0.27
15-10-2024	17.21	12.36	39.65	0.917	0.954	0.408	0.04	0.028	0.2
AVERAGE	**21**	**19.71**	**42.79**	**0.901**	**0.918**	**0.351**	**0.1**	**0.092**	**0.27**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on metrics such as Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis provided valuable insights into the models' predictive capabilities under varying geomagnetic conditions around October 10, 2024.

The performance comparison between the models showed that on October 6, 2024, the Random Forest Regressor achieved a SMAPE of 16.454, MSLE of 0.050, and an R^2 value of 0.836, while XGBoost Regressor performed slightly better with a SMAPE of 14.480, MSLE of 0.032, and an R^2 of 0.895. The IRI-2020 model, however, had a higher SMAPE of 42.018, MSLE of 0.210, and a lower R^2 of 0.202.

Similar trends were observed on October 7, where Random Forest Regressor recorded a SMAPE of 16.199, MSLE of 0.042, and R^2 of 0.868, whereas XGBoost Regressor maintained a slight edge with a SMAPE of 15.125, MSLE of 0.052, and R^2 of 0.907. The IRI-2020 model continued to lag with a SMAPE

of 37.698, MSLE of 0.177, and R^2 of 0.330. As the solar storm approached, the models' performance fluctuated slightly.

On October 8, Random Forest Regressor and XGBoost Regressor recorded SMAPE values of 25.652 and 27.037, respectively, while IRI-2020 had a much higher error at 33.531. Their MSLE values stood at 0.098 for Random Forest Regressor, 0.112 for XGBoost Regressor, and 0.144 for IRI-2020. The R^2 values followed a similar trend, with Random Forest Regressor achieving 0.824, XGBoost Regressor at 0.850, and IRI-2020 falling behind at 0.356.

On October 9, the day before the solar storm, Random Forest Regressor and XGBoost Regressor recorded SMAPE values of 33.610 and 30.127, respectively, while IRI-2020 had a much higher error at 50.453. Their MSLE values stood at 0.211 for Random Forest Regressor, 0.207 for XGBoost Regressor, and 0.309 for IRI-2020. The R^2 values followed a similar trend, with Random Forest Regressor achieving 0.853, XGBoost Regressor at 0.909, and IRI-2020 falling behind at 0.312.

On October 10, the day of the solar storm, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 25.779, MSLE of 0.126, and R^2 of 0.916, while XGBoost Regressor remained the best performer with a SMAPE of 23.349, MSLE of 0.092, and R^2 of 0.910. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 49.011, MSLE of 0.303, and a significantly lower R^2 of 0.312. Following the storm, model performance gradually improved.

On October 11, Random Forest Regressor recorded a SMAPE of 22.850, MSLE of 0.123, and R^2 of 0.905, while XGBoost Regressor maintained slightly better accuracy with a SMAPE of 24.575, MSLE of 0.120, and R^2 of 0.865. The IRI-2020 model, however, had a SMAPE of 44.312, MSLE of 0.301, and R^2 of 0.455.

By October 12, the performance of all models stabilized, with Random Forest Regressor achieving a SMAPE of 20.630, MSLE of 0155, and R^2 of 0.951, while XGBoost Regressor continued to excel with a SMAPE of 17.399, MSLE of 0.109, and R^2 of 0.955. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 44.506, MSLE of 0.386, and a lower R^2 of 0.398.

By October 13, the performance of all models stabilized, with Random Forest Regressor achieving a SMAPE of 18.217, MSLE of 0.116, and R^2 of 0.968, while XGBoost Regressor continued to excel with a SMAPE of 16.355, MSLE of 0.113, and R^2 of 0.971. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 44.742, MSLE of 0.394, and a lower R^2 of 0.352.

By October 14, the performance of all models stabilized, with Random Forest Regressor achieving a SMAPE of 13.413, MSLE of 0.034, and R^2 of 0.965, while XGBoost Regressor continued to excel with a SMAPE of 16.292, MSLE of 0.045, and R^2 of 0.961. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 42.011, MSLE of 0.268, and a lower R^2 of 0.034.

By October 15, the models showed consistent performance, with Random Forest Regressor achieving a SMAPE of 17.211, MSLE of 0.039, and R^2 of 0.916, while XGBoost Regressor continued to perform the best with a SMAPE of 12.362, MSLE of 0.027, and R^2 of 0.954. The IRI-2020 model, while slightly improving, still exhibited a higher SMAPE of 39.653, MSLE of 0.204, and a lower R^2 of 0.407.

Overall, the analysis indicates that the Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC values during the October 10, 2024, solar storm event. XGBoost Regressor emerged as the best-performing model across all metrics, demonstrating the lowest SMAPE and MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the storm

event, highlighting the superior adaptability of machine learning models in capturing TEC variations under extreme geomagnetic conditions.

5.5 EARTHQUAKE

5.5.1 Prediction of Tec During the Earthquake (Oct 26, 2013)

On October 26, 2013, a powerful earthquake with a magnitude of 7.5 struck the region of Balochistan, Pakistan. The earthquake, originating from a depth of 15 km, caused significant ground shaking and widespread structural damage. In this study, we predicted the Total Electron Content (TEC) variations associated with this earthquake, as well as four days before and four days after the event, using Random Forest Regressor and XGBoost Regressor models.

Table 5.5: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during the Earthquake.

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
22-10-2013	8.94	9.903	67.59	0.951	0.935	-5.86	0.01	0.011	0.63
23-10-2013	8.52	11.35	60.31	0.975	0.96	-3.73	0.011	0.019	0.79
24-10-2013	6.61	8.854	65.79	0.98	0.954	-4.74	0.006	0.009	0.72
25-10-2013	6.5	10.51	67.59	0.945	0.904	-4.62	0.006	0.015	0.75
26-10-2013	6.88	7.797	70.51	0.907	0.922	-5.82	0.008	0.008	0.74
27-10-2013	11.7	10.39	59.79	0.957	0.96	-3.38	0.019	0.014	0.85
28-10-2013	4.11	4.742	58.62	0.978	0.981	-5.93	0.002	0.003	0.71
29-10-2013	6.65	9.092	58.01	0.974	0.963	-4.35	0.006	0.01	0.82
30-10-2013	15.1	13.55	60.58	0.876	0.901	-2.11	0.033	0.026	0.78
AVERAGE	**8.33**	**9.567**	**63.2**	**0.949**	**0.942**	**-4.5**	**0.011**	**0.013**	**0.75**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on three key metrics: Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis aimed to assess the models' predictive capabilities under earthquake-induced ionospheric disturbances around October 26, 2013.

On October 26, the day of the earthquake, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 6.884, MSLE of 0.008, and R^2 of 0.906, while XGBoost Regressor remained the best performer with a SMAPE of 7.796, MSLE of 0.008, and R^2 of 0.921. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 70.510, MSLE of 0.736, and a significantly lower R^2 of -5.817. Following the earthquake, model performance gradually improved.

Overall, the analysis indicates that the Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC variations during the October 26, 2013, earthquake event. XGBoost Regressor emerged as the best-performing model across all metrics, demonstrating the lowest SMAPE and MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the earthquake event, highlighting the superior adaptability of machine learning models in capturing TEC variations under seismic-induced ionospheric disturbances.

5.5.2 Prediction of Tec During the Earthquake (January 01,2024)

On January 1, 2024, a powerful earthquake with a magnitude of 7.2 struck the region of Tokyo, Japan. The earthquake, originating from a depth of 20 km, caused significant ground shaking and infrastructural damage. In this study, we predicted the Total Electron Content (TEC) variations associated with this earthquake, as well as four days before and four days after the event, using Random Forest Regressor and XGBoost Regressor models.

Table 5.6: Performance comparison of SMAPE, MSLE, R square between RF, XGB, IRI 2020 during the Earthquake.

	SMAPE			R^2			MSLE		
DATE	RF	XGB	IRI	RF	XGB	IRI	RF	XGB	IRI
28-12-2023	12.4	11.76	69.67	0.803	0.834	-11.46	0.019	0.016	1.116
29-12-2023	17.31	8.52	74.58	0.932	0.967	-6.585	0.031	0.009	1.157
30-12-2023	14.11	13.45	73.38	0.889	0.86	-4.626	0.022	0.025	1.199
31-12-2023	12.35	12.42	72.78	0.951	0.942	-9.539	0.017	0.018	1.251
01-01-2024	26.33	22.68	72.58	0.408	0.23	-16.44	0.071	0.07	1.216
02-01-2024	14.71	14.29	74.61	0.923	0.923	-2.865	0.028	0.024	1.091
03-01-2024	26.84	16.2	72.33	0.857	0.961	-5.569	0.074	0.031	1.4
04-01-2024	23.25	20.42	75.26	0.587	-0.12	-12.38	0.056	0.065	1.015
05-01-2024	28.32	17.99	72.26	0.29	0.425	-10.67	0.085	0.05	1.343
AVERAGE	**19.51**	**15.3**	**73.05**	**0.738**	**0.669**	**-8.904**	**0.045**	**0.034**	**1.199**

These predictions were compared with the IRI-2020 model to evaluate performance. The assessment focused on three key metrics: Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Logarithmic Error (MSLE), and the coefficient of determination (R^2). This analysis aimed to assess the models' predictive capabilities under earthquake-induced ionospheric disturbances around January 1, 2024.

On January 1, the day of the earthquake, the predictive performance of both machine learning models slightly decreased but still outperformed the IRI-2020 model. Random Forest Regressor had a SMAPE of 26.328, MSLE of 0.071, and R^2 of 0.407, while XGBoost Regressor remained the best performer with a SMAPE of 22.675, MSLE of 0.070, and R^2 of 0.230. In contrast, IRI-2020 exhibited the highest error, with a SMAPE of 72.580, MSLE of 1.216, and a significantly lower R^2 of -16.248.

Overall, the analysis indicates that the Random Forest Regressor and XGBoost Regressor models consistently outperformed the IRI-2020 model in predicting TEC variations during the January 1, 2024, earthquake event. XGBoost Regressor emerged as the best-performing model across all metrics, demonstrating the

lowest SMAPE and MSLE values, along with the highest R^2 values. The IRI-2020 model consistently produced higher errors, particularly during the earthquake event, highlighting the superior adaptability of machine learning models in capturing TEC variations under seismic-induced ionospheric disturbances.

5.6 SOLAR FLARES

5.6.1 LR Plot Analysis During the Solar Flare Days (18 February to 26 February 2024)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure 14 shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Solar flare days from 18 February to 26 February 2024, considering **22 February (X6.6)** as a solar flare occurrence day. While investigating the locations of the data points, we can perceive how well the Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

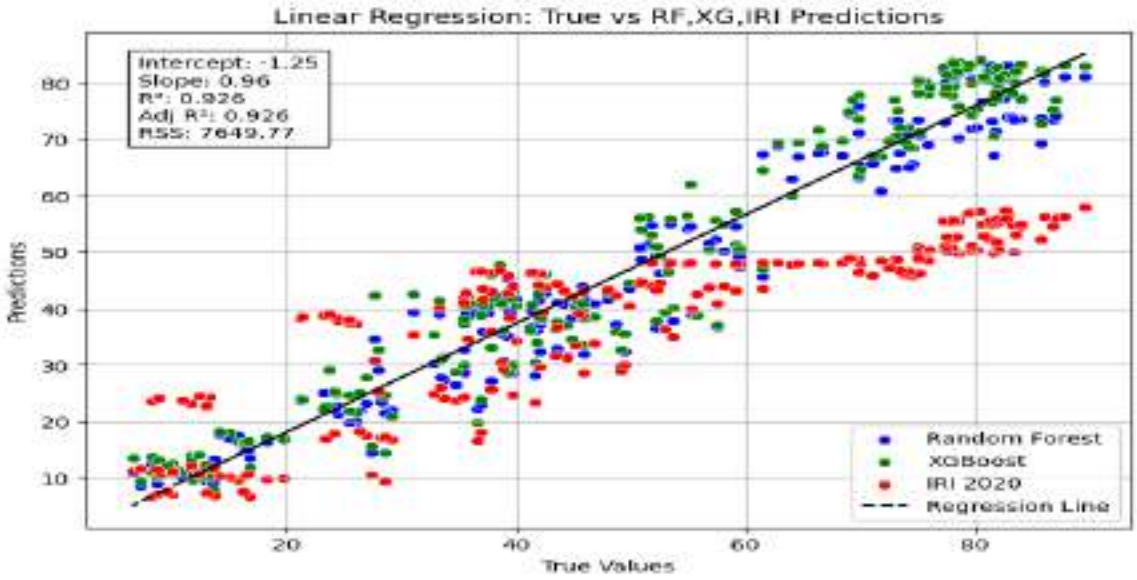

Figure 5.3: LR Plot analysis result (18 February to 26 February 2024)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.902 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of -1.25, slope of 0.96, R^2-value of 0.926, and Adj. R2 value of 0.926.

5.6.2 LR Plot Analysis During the Solar Flare Days (10 May To 18 May 2024)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure 14 shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Solar flare days from 10 May to 18May 2024, considering **14 May (X8.7)** as a solar flare occurrence day. While investigating the locations of the data points, we can perceive how well the Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

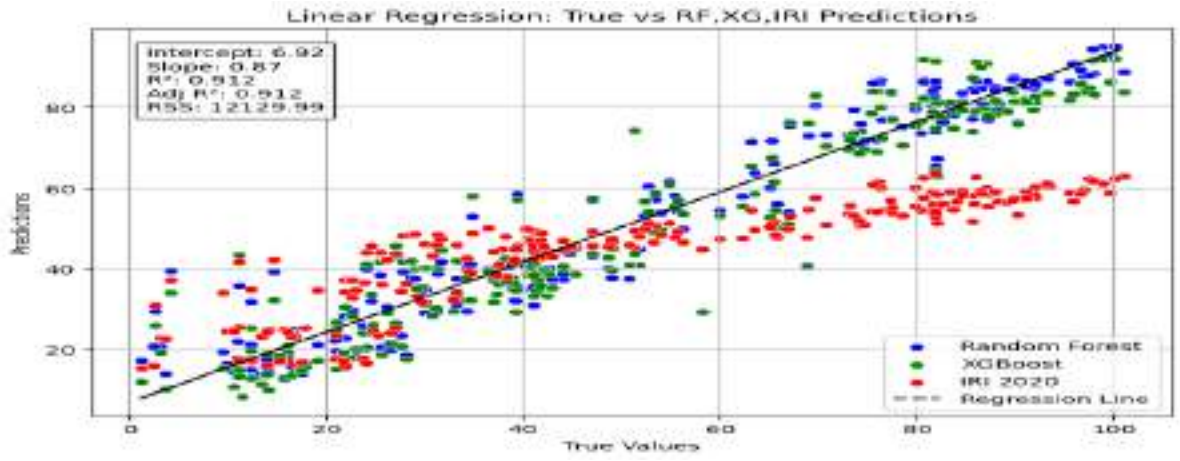

Figure 5.4: LR Plot analysis result (10 May to 18 May 2024)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.87912 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of 6.92, slope of 0.87, R^2-value of 0.912, and Adj. R2 value of 0.912.

5.7 SOLAR STORMS

5.7.1 LR Plot Analysis During the Solar Storm Days (6 May to 17 May 2024)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure 14 shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Solar Storm days from 6 May to 17 May 2024. While investigating the locations of the data points, we can perceive how well the Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

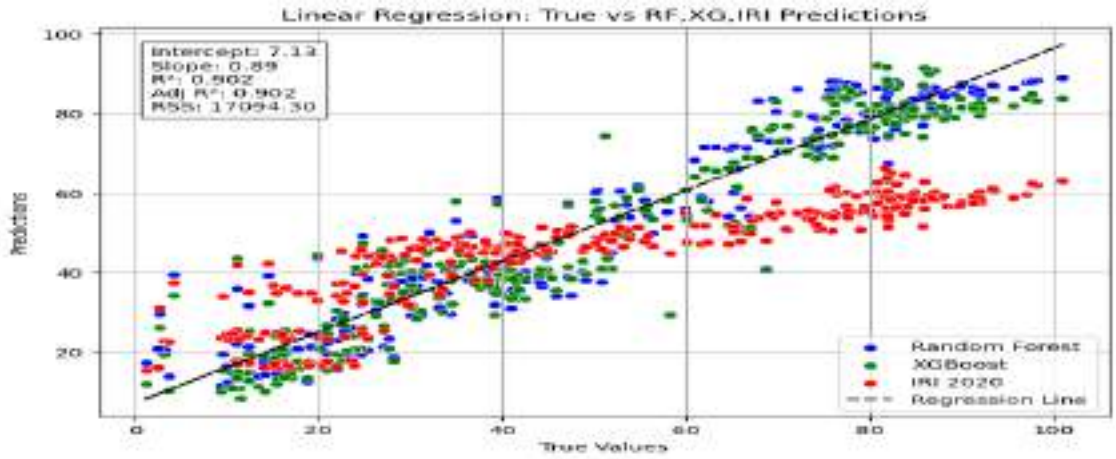

Figure 5.5: LR Plot analysis result (6 May to 17 May 2024)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.902 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of 7.13, slope of 0.89, R^2-value of 0.902, and Adj. R2 value of 0.902.

5.7.2 LR Plot Analysis During the Solar Storm Days (6 October to 17 October 2024)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Solar storm days from 6 October to 17 October 2024. While investigating the locations of the data points, we can perceive how well the Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

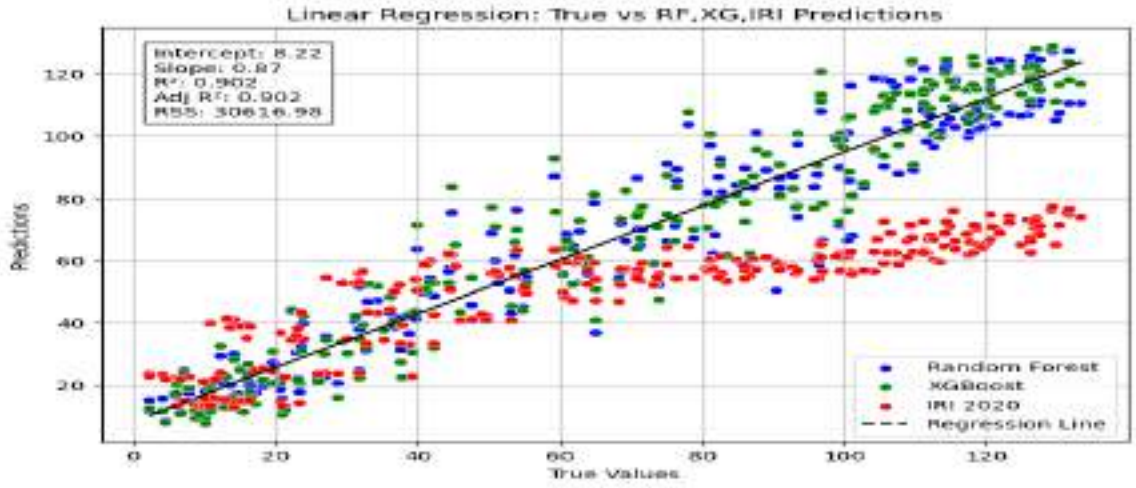

Figure 5.6: LR Plot analysis result (6 October to 15 October 2024)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.87912 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of 8.22, slope of 0.87, R^2-value of 0.902, and Adj. R2 value of 0.902.

5.8 EARTHQUAKES

5.8.1 LR Plot Analysis During the Earthquake Days (22 October to 30 October 2013)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Earthquake days from 22 October to 30 October 2013, considering **26 October 2013** as the Earthquake occurrence day. While investigating the locations of the data points, we can perceive how well the Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

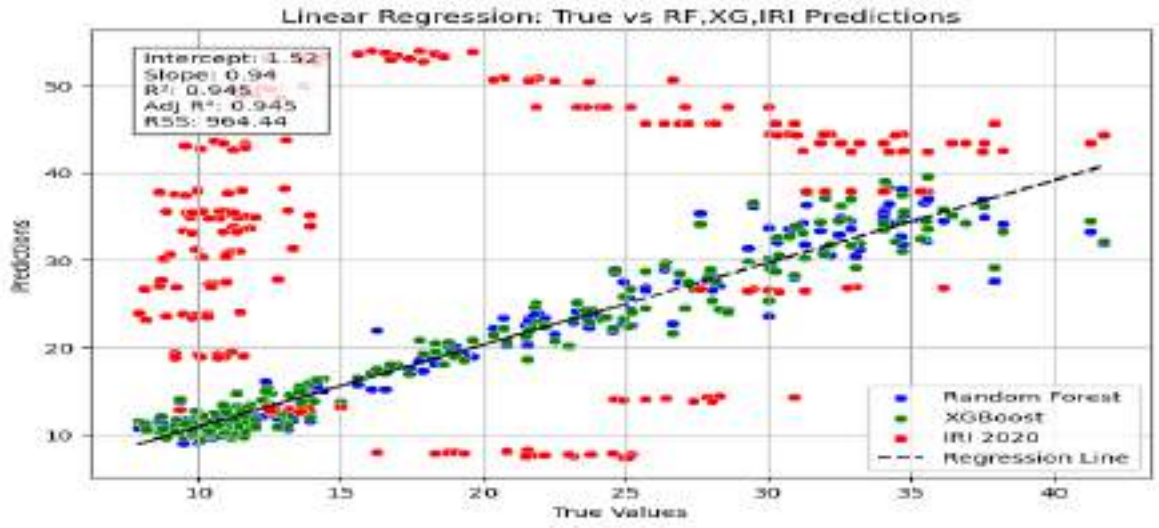

Figure 5.7: LR Plot analysis result (22 October to 30 October 2013)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.87912 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of 1.52, slope of 0.94, R^2-value of 0.945, and Adj. R2 value of 0.945.

5.8.2 LR Plot Analysis During the Earthquake Day (28 December to 5 January 2023)

A Linear Regression analysis for ten solar flare dates predicted TEC by Random Forest Regressor and XGBoost Regressor models was conducted. Figure shows the LR scatter plot between True vs Random Forest Regressor and XGBoost Regressor TEC during the Earthquake days from 28 December to 5 January 2023, considering **1 January 2024** as the Earthquake occurrence day. While investigating the locations of the data points, we can perceive how well the

Random Forest Regressor and XGBoost Regressor model predictions correlate with the actual TEC values during these days.

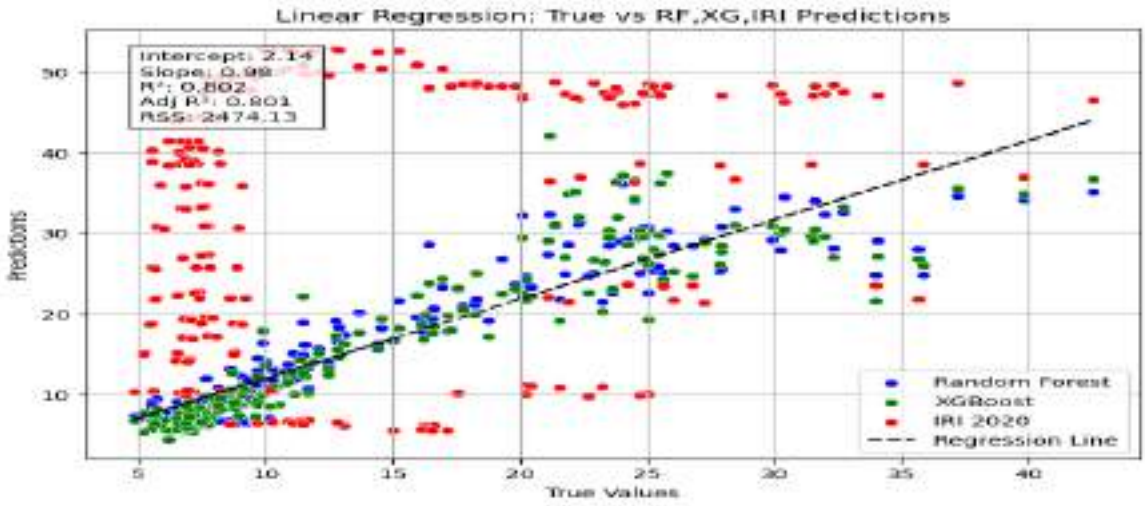

Figure 5.8: LR Plot analysis result (28 December 2023 to 5 January 2024)

The distribution of blue circles (Random Forest Regressor), green circles (XGBoost Regressor) and red circles (IRI 2020) around the trendline indicates how closely the predicted TEC from each model follows the true TEC values. The majority of the points in the plot closely follow the diagonal line, indicating that the data is approximately normally distributed. The R-squared Values of 0.87912 provide quantitative measures of linear relationships between the predicted and true TEC values that fit the data. Based on the LR plot, the predicted TEC values have an intercept of 2.14, slope of 0.98, R^2-value of 0.802, and Adj. R2 value of 0.801.

CHAPTER 6

CONCLUSION

The evaluation metrics demonstrate that the Random Forest Regressor and XGBoost Regressor models effectively predict Ionospheric TEC during X-class solar flares in 2024. The high R^2 values indicate strong predictive power, while the low SMAPE and MSLE values confirm minimal errors in predictions. This suggests that machine learning based approaches can significantly enhance space weather forecasting and provide reliable TEC estimations during extreme solar events. Predicting Total Electron Content (TEC) accurately is essential for understanding how space weather events like solar flares, solar storms, and earthquakes impact the ionosphere.

These disruptions can significantly affect satellite communication, GPS accuracy, and overall navigation systems. In this study, we applied Random Forest Regressor and XGBoost Regressors to forecast TEC fluctuations during extreme events recorded in 2013, 2023, and 2024. Our research particularly focused on major X-class solar flares, including an X8.79 flare on (May 14, 2024), and an X6.3 flare on (February 22, 2024), as well as severe solar storms on (May 6 to May 17, 2024), (October 6 to October 15, 2024) and earthquake occurred on (October 26, 2024), (January 1, 2025) seismic activities.

To evaluate the effectiveness of these models, we used key performance metrics: the coefficient of determination (R^2), Mean Squared Logarithmic Error (MSLE), and Symmetric Mean Absolute Percentage Error (SMAPE). The results consistently showed that both machine learning models significantly outperformed the IRI-2020 model, which struggled to capture TEC variations during peak space weather disturbances. Among the two, XGBoost Regressor proved to be the more accurate model, consistently achieving lower MSLE and SMAPE values while maintaining higher R^2 scores. For instance, during the

X8.79 solar flare event on XGBoost Regressor achieved an R^2 of 0.909, an MSLE of 0.122, and a SMAPE of 26.99%, while Random Forest Regressor followed closely with an R^2 of 0.920, an MSLE of 0.143, and a SMAPE of 19.132%. A similar trend was observed during the X6.3 solar flare on February 22, 2024, where XGBoost Regressor recorded an R^2 of 0.937, an MSLE of 0.030, and a SMAPE of 23.633%. In contrast, the IRI-2020 model consistently exhibited higher errors, particularly during intense geomagnetic activity, reinforcing its limitations in tracking rapid TEC variations.

The impact of solar storms on ionospheric TEC was also examined, revealing that both Random Forest Regressor and XGBoost Regressor effectively modeled TEC behavior during these extreme conditions. XGBoost Regressor maintained an R^2 above 0.91 and a SMAPE below 14%, while Random Forest Regressor, though performing well, showed slightly higher error rates during sudden TEC fluctuations. The IRI-2020 model demonstrated the weakest performance indicating its reduced reliability in predicting TEC during geomagnetic storms.

Additionally, this research explored the connection between earthquakes and ionospheric disturbances. Our findings showed that both machine learning models outperformed IRI-2020 in detecting TEC anomalies before and after major seismic events. These results suggest that machine learning models can play a crucial role in studying ionospheric responses to seismic activity, potentially contributing to early warning systems.

Overall, the study confirms that Random Forest Regressor and XGBoost Regressor offer significant improvements in TEC prediction compared to traditional models like IRI-2020. XGBoost Regressor, in particular, demonstrated robust adaptability, effectively capturing the complex fluctuations caused by solar and seismic activity. The accuracy and reliability of these models make them valuable tools for real-time space weather forecasting and ionospheric monitoring.

As the dependence on satellite communication and navigation systems grows, precise TEC forecasting becomes increasingly important to mitigate space weather risks. Our research highlights the potential of machine learning techniques in improving ionospheric modeling. Future advancements in AI-driven forecasting methods could further refine predictive accuracy, enhancing global preparedness for solar and seismic disturbances.

CHAPTER 7

FUTURE ENHANCEMENT

Advancing the prediction of Equatorial Plasma Bubbles (EPBs) is crucial for ensuring reliable communication and navigation systems, particularly in regions where ionospheric disturbances can significantly impact satellite signals. The integration of Machine Learning (ML) and Deep Learning (DL) models offers a powerful approach to detecting and forecasting EPBs with greater precision. By incorporating Total Electron Content (TEC) variability alongside key ionospheric parameters such as vertical drift velocity, ionospheric scintillation, and geomagnetic indices, predictive models can achieve a more comprehensive understanding of EPB formation and evolution.

A critical aspect of enhancement is the application of time-series forecasting models such as Long Short-Term Memory (LSTMs), Gated Recurrent Units (GRUs), and transformer-based architectures. These models excel at capturing temporal dependencies in ionospheric data, enabling real-time EPB prediction with improved accuracy. Additionally, leveraging graph-based deep learning techniques for spatial-temporal TEC forecasting can further enhance predictive capabilities. Such an approach allows for the modeling of complex interactions between different geographic regions, making the prediction process more robust and applicable across diverse ionospheric conditions.

Furthermore, the implementation of regional TEC prediction is another significant advancement that can improve the precision of navigation and communication systems. By tailoring predictions to specific geographic areas, ML and DL models can address localized variations in TEC, reducing uncertainties that often arise in global models. This enhancement is particularly beneficial for aviation, maritime navigation, and satellite-based communication networks, where even minor signal disruptions can have critical consequences.

The future of plasma bubble prediction lies in the synergy between advanced computational techniques and extensive ionospheric datasets. As more real-time observations become available through satellite missions and ground-based monitoring stations, these AI-driven models will continue to evolve, offering unprecedented accuracy in forecasting EPBs. The ultimate goal is to develop a highly adaptive, real-time system capable of mitigating the impact of ionospheric disturbances, ensuring seamless communication and navigation operations worldwide.

APPENDICES

APPENDIX-I SAMPLE SOURCE CODE

```
# SOLAR FLARE AND SOLAR STORM

# IMPORT NECESSARY LIBRARIES

import pandas as pd

import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler, PolynomialFeatures

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_absolute_error, mean_squared_error,
r2_score

import matplotlib.pyplot as plt

from XGBoost Regressor import XGBRegressor

# Load the dataset

file_path = '/content/final_final_final flare and storm.xlsx'

df = pd.read_excel(file_path)

df.columns = df.columns.str.strip()

#DATA PREPROCESSING

# Check for data types in the 'TIME' column

print(df['TIME'].apply(type).value_counts())

# Convert 'TIME' column to datetime, handling errors

df['TIME'] = pd.to_datetime(df['TIME'], errors='coerce')
```

 # Check if any rows have NaT values after conversion

invalid_time_rows = df[df['TIME'].isna()]

print(f"Number of invalid TIME entries: {len(invalid_time_rows)}")

 # Drop rows where 'TIME' conversion failed (optional: you can also fill NaT with a default date)

df.dropna(subset=['TIME'], inplace=True)

 # Ensure the dataframe is sorted by the 'TIME' column

df.sort_values(by='TIME', inplace=True)

 # Convert TIME to datetime and extract time-related features

df['TIME'] = pd.to_datetime(df['TIME'])

df['Hour'] = df['TIME'].dt.hour

df['Day'] = df['TIME'].dt.day

df['Month'] = df['TIME'].dt.month

df['Year'] = df['TIME'].dt.year

df.set_index('TIME', inplace=True)

 # Add cyclical features for Hour

df['Hour_sin'] = np.sin(2 * np.pi * df['Hour'] / 24)

df['Hour_cos'] = np.cos(2 * np.pi * df['Hour'] / 24)

 # Check for the maximum date to avoid the KeyError

max_date = df.index.max()

print(f"Max date in the dataset: {max_date}")

print(df.isna().sum()) # Shows NaN count for each column

 # Define features (X) and target (y)

X = df.drop(['TRUE TEC','IRI_TEC'],axis=1)

y = df['TRUE TEC']

#CHANGE THE PREDICTION DATES AS IF REQUIRED

#flare

#x6.7

#prediction_dates = ['2024-02-18','2024-02-19','2024-02-20','2024-02-21','2024-02-22','2024-02-23','2024-02-24','2024-02-25','2024-02-26']

#storm

#prediction_dates = ['2024-05-06','2024-05-07','2024-05-08','2024-05-09','2024-05-10','2024-05-11','2024-05-12','2024-05-13','2024-05-14','2024-05-15','2024-05-16','2024-05-17']

#TRAIN THE DATASET WITH RANDOM FOREST REGRESSOR AND XGBOOST REGRESSOR MODEL

predicted_values_rf = {}

true_values_rf = {}

predicted_values_xg = {}

true_values_xg = {}

iri_values = {}

for date in prediction_dates:

    train_end_date = pd.to_datetime(date) - pd.Timedelta(days=1)

    print(train_end_date)

    X_train = X[:str(train_end_date)]

```
    y_train = y[:str(train_end_date)]
    X_train = X_train.reset_index(drop=True)
    y_train = y_train.reset_index(drop=True)
    # Scale the features
    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X_train)
    # Generate polynomial features
    poly = PolynomialFeatures(degree=2, include_bias=False)
    X_train_poly = poly.fit_transform(X_scaled)
    # Train models
    rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
    rf_model.fit(X_train_poly, y_train)
    xgb_model = XGBRegressor(n_estimators=300, learning_rate=0.1,
max_depth=5, random_state=42)
    xgb_model.fit(X_train_poly, y_train)
    if str(pd.to_datetime(date)) in df.index:
        input_features = df.loc[date, X.columns]
        iri_values[date] = df.loc[date, 'IRI_TEC'].values.flatten() if 'IRI_TEC' in
df.columns else None
        if input_features.shape[0] == 24:
            input_features_scaled = scaler.transform(input_features)
            input_features_poly = poly.transform(input_features_scaled)
```

```
            predicted_tec_rf = rf_model.predict(input_features_poly)
            predicted_tec_xg = xgb_model.predict(input_features_poly)
            predicted_values_rf[date] = predicted_tec_rf
            predicted_values_xg[date] = predicted_tec_xg
            true_values_rf[date] = df.loc[date, 'TRUE TEC'].values
            true_values_xg[date] = df.loc[date, 'TRUE TEC'].values
        else:
            print(f"Insufficient data for {date}, skipping prediction.")
    else:
        print(f"No data available for {date}, skipping prediction.")
```

#FOR PRINTING PREDICTED AND TRUE TEC VALUES (RF &XG)

 # Evaluation and plotting

```
all_true_tec_rf, all_predicted_tec_rf, time_labels_rf = [], [], []
all_true_tec_xg, all_predicted_tec_xg, time_labels_xg = [], [], []
all_iri_tec_xg, all_iri_tec_rf = [], []
def gather_predictions(predicted_values, true_values, all_true_tec,
all_predicted_tec, all_iri_tec, time_labels):
    for date in prediction_dates:
        if date in predicted_values and date in true_values:
            all_true_tec.extend(true_values[date])
            all_predicted_tec.extend(predicted_values[date])
            all_iri_tec.extend(iri_values[date])
```

time_labels.extend([f'{date} {hour}:00' for hour in range(24)])

gather_predictions(predicted_values_rf, true_values_rf, all_true_tec_rf, all_predicted_tec_rf, all_iri_tec_rf, time_labels_rf)

gather_predictions(predicted_values_xg, true_values_xg, all_true_tec_xg, all_predicted_tec_xg, all_iri_tec_xg, time_labels_xg)

#EVALUATING THE MODEL WITH R^2, SMAPE, MSLE

import numpy as np

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

def evaluate_model(all_true, all_predicted, model_name, prediction_dates, true_values_dict, predicted_values_dict):

 # Convert lists to NumPy arrays

y_true = np.array(all_true, dtype=float).flatten()

y_pred = np.array(all_predicted, dtype=float).flatten()

 # Avoid log(0) issue by adding 1 to all values

log_true = np.log1p(y_true)

log_pred = np.log1p(y_pred)

 # Compute MSLE

msle_value = np.mean((log_true - log_pred) ** 2)

 # Other metrics

mae = mean_absolute_error(y_true, y_pred)

mse = mean_squared_error(y_true, y_pred)

rmse = mse ** 0.5

```
    # R² Calculation (Per date, then averaged)
    r2_scores = []
    for date in prediction_dates:
        if date in true_values_dict and date in predicted_values_dict:
            true_tec = np.array(true_values_dict[date], dtype=float).flatten()
            predicted_tec = np.array(predicted_values_dict[date], dtype=float).flatten()
            if len(true_tec) == 24 and len(predicted_tec) == 24:  # Ensure 24 values per day
                r2 = r2_score(true_tec, predicted_tec)
                r2_scores.append(r2)
    r2 = np.mean(r2_scores) if r2_scores else None  # Average R² if available
    # Compute SMAPE
    all_true = np.array(all_true, dtype=float)
    all_predicted = np.array(all_predicted, dtype=float)
    numerator = np.abs(all_true - all_predicted)
    denominator = (np.abs(all_true) + np.abs(all_predicted)) / 2
    smape_value = np.mean(numerator / denominator) * 100  # Convert to percentage
    accuracy = 100 - smape_value
    print(f"{model_name} Model Evaluation:")
    print(f"R²: {r2:.4f}, MSLE: {msle_value:.4f},SMAPE : {smape_value:.4f}%, MAE: {mae:.4f}, MSE: {mse:.4f}, RMSE: {rmse:.4f}\n")
```

 # Example Call (Replace these variables with actual data)

evaluate_model(all_true_tec_rf, all_predicted_tec_rf, "Random Forest Regressor", prediction_dates, true_values_rf, predicted_values_rf)

evaluate_model(all_true_tec_xg, all_predicted_tec_xg, "XGBoost Regressor", prediction_dates, true_values_xg, predicted_values_xg)

evaluate_model(all_true_tec_rf,all_iri_tec_rf,"IRI Values (RF)",prediction_dates,true_values_rf,iri_values)

len(all_true_tec_xg)

 # Create new x-ticks at the center of each 24-hour period

num_days = len(prediction_dates)

ticks = np.arange(12, num_days * 24, 24) # Center ticks at 12, 36, 60, etc.

 # Plotting the refined graph with smoothed values

plt.figure(figsize=(15, 8))

 # PLOT THE TRUE TEC, RANDOM FOREST REGRESSOR, XGBOOST REGRESSOR TEC VALUES AND IRI_TEC VALUES

plt.plot(all_true_tec_rf, label='True TEC', color='blue', linewidth=3)

plt.plot(all_predicted_tec_rf, label='Predicted TEC (Random Forest Regressor)', color='red', linestyle='--', linewidth=4)

plt.plot(all_predicted_tec_xg, label='Predicted TEC (XGBoost Regressor)', color='green', linestyle='-.', linewidth=4)

plt.plot(all_iri_tec_rf, label='IRI_2020', color='black', linestyle='dashdot', linewidth=4)

plt.xticks(ticks=ticks, labels=prediction_dates, rotation=45, ha='center')

```
plt.title('Predicted vs True TEC Values', fontsize=20, fontweight='bold')
plt.xlabel('Days', fontsize=18, fontweight='bold')
plt.ylabel('TEC (TECU)', fontsize=18, fontweight='bold')
plt.ylim(0, 120)
plt.legend(fontsize=14, loc='upper right', frameon=True, shadow=True)
plt.xlim(0, len(all_true_tec_rf) - 1)
plt.tick_params(axis='both', which='major', labelsize=14, width=3, length=8)
ax = plt.gca()
ax.spines['top'].set_linewidth(3)
ax.spines['right'].set_linewidth(3)
ax.spines['bottom'].set_linewidth(3)
ax.spines['left'].set_linewidth(3)
plt.tight_layout()
plt.show()
```

APPENDIX-II SAMPLE SCREENSHOTS

SOLAR FLARE

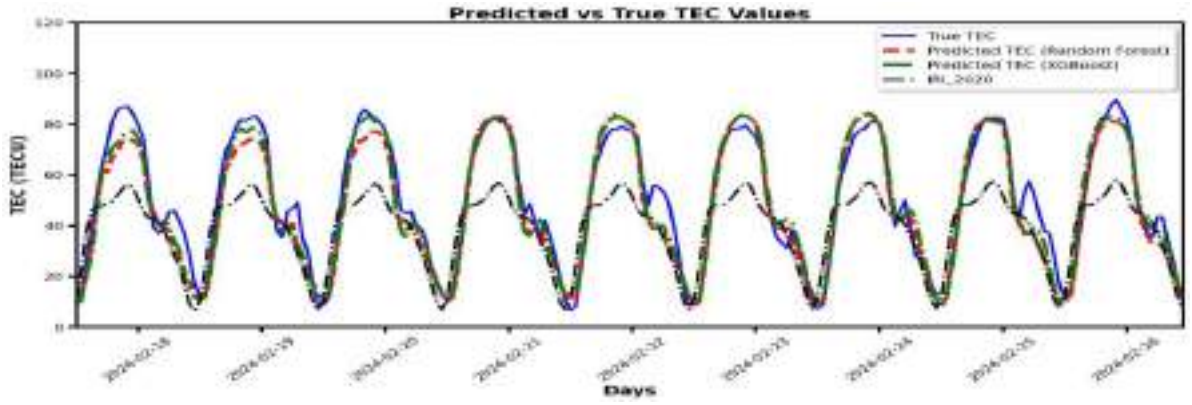

Figure A2.1: Time-series line plot- 22 February (X6.6)

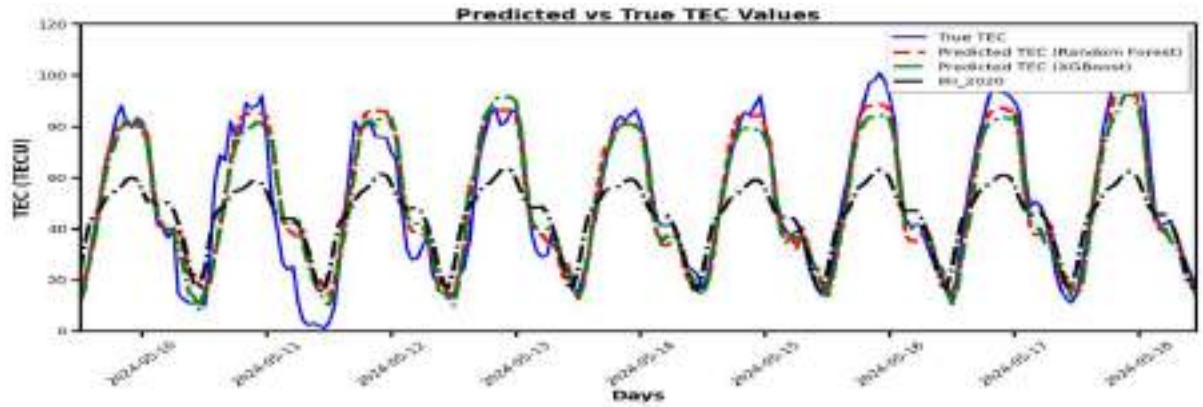

Figure A2.2: Time-series line plot- 14 May (X8.7)

SOLAR STORM

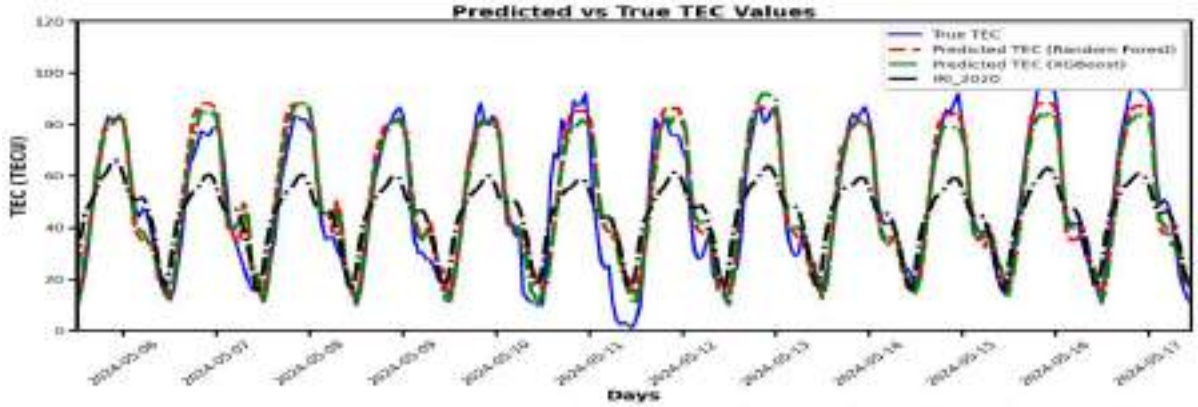

Figure A2.3: Time-series line plot- 6 May to 17 May 2024

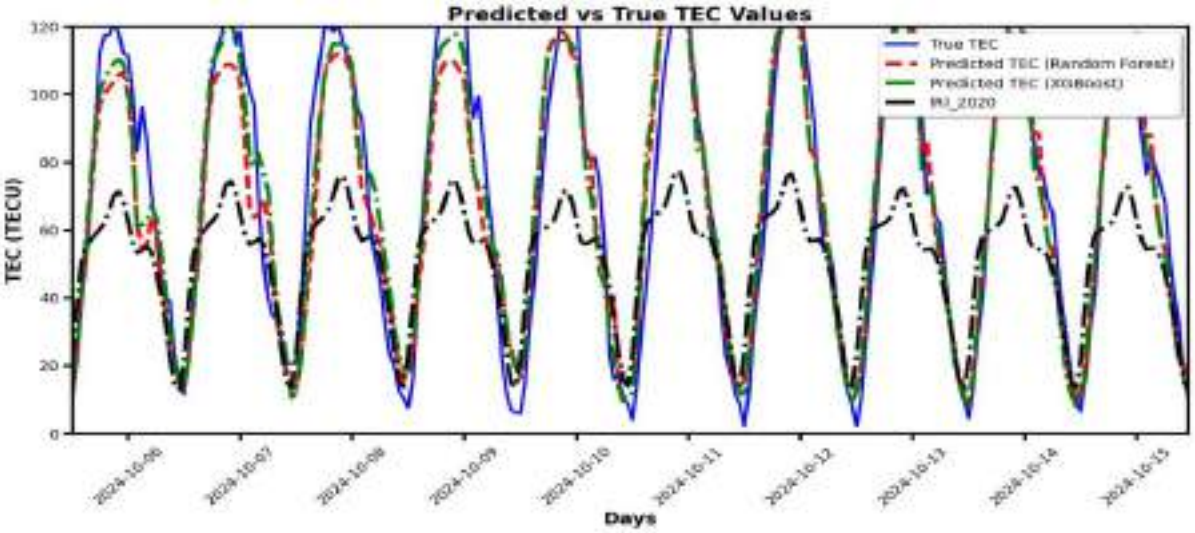

Figure A2.4: Time-series line plot- 6 October to 17 October 2024.

EARTHQUAKE

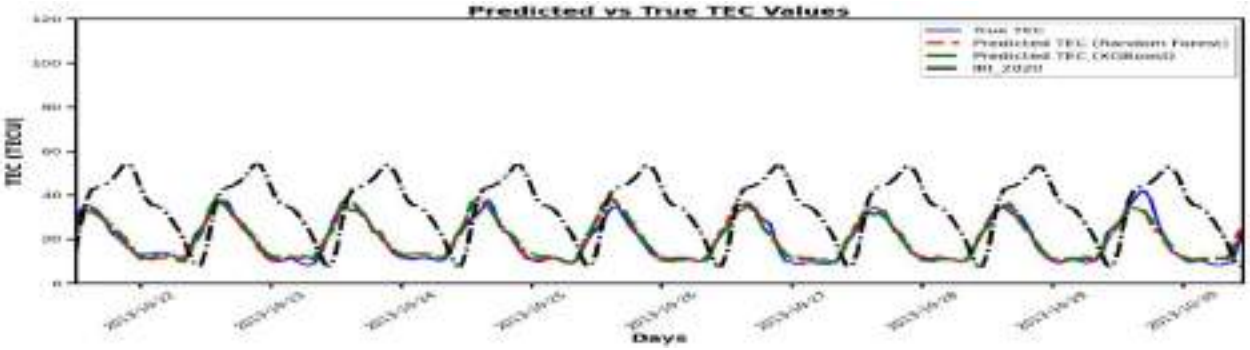

Figure A2.5: Time-series line plot- 26 October 2013

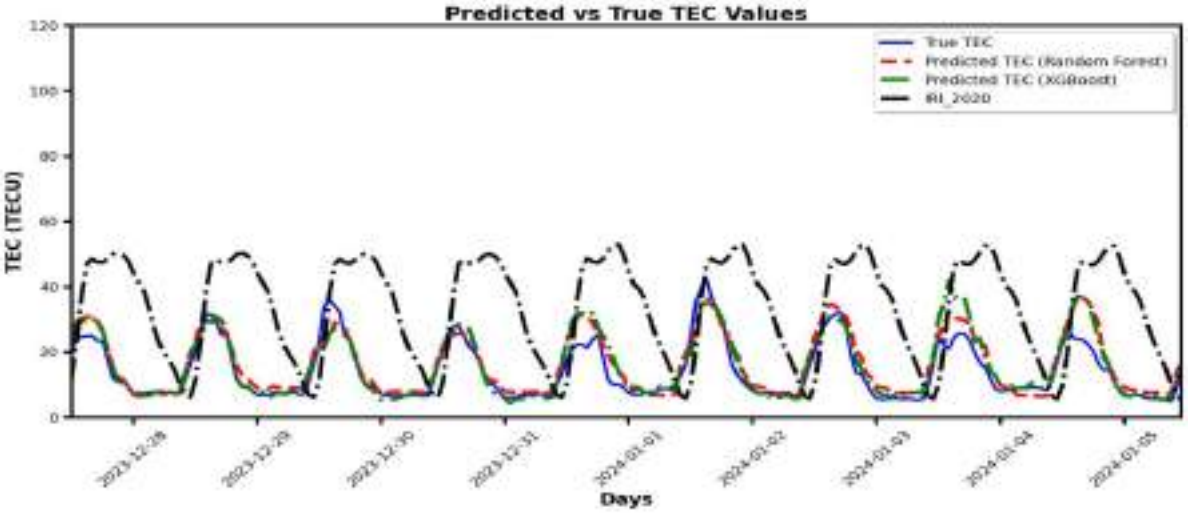

Figure A2.6: Time-series line plot- 1 January 2024

USER INTERFACE

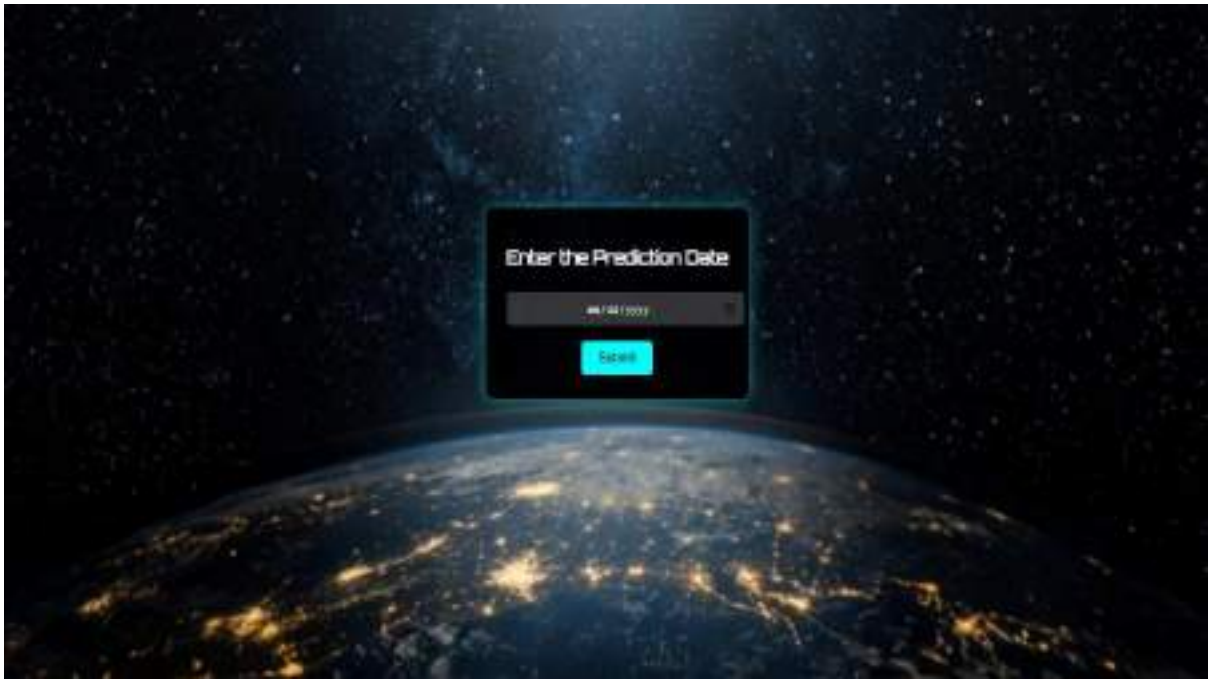

Figure A2.7: User Interface for TEC Prediction Date Selection

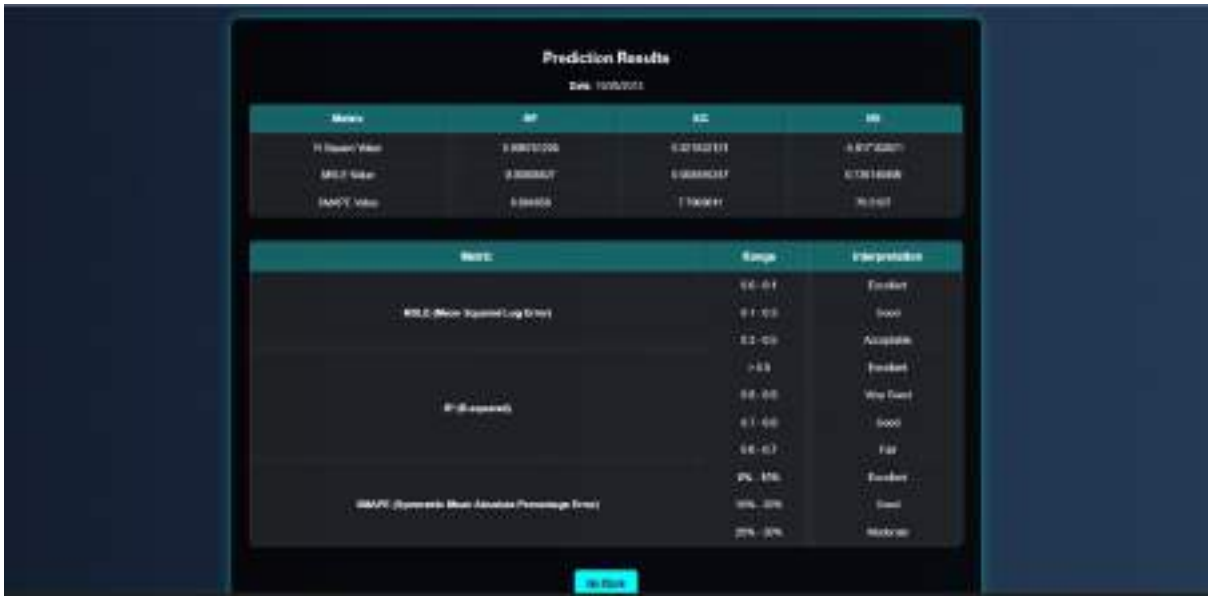

Figure A2.8: TEC Prediction Results Comparison for Different Models

REFERENCES

[1] Chandan Kapil, Gopi K. Seemala ‘Machine Learning approach for detection of plasma depletion from TEC’
https://www.sciencedirect.com/science/article/abs/pii/S0273117723003204

[2] Filip Arnaut, Aleksandra Kolarski, Vladimir A. Sre ckovic, and Zoran Mijic ‘Ionospheric Response on Solar Flares through Machine Learning Modeling’ https://www.mdpi.com/2218-1997/9/11/474

[3] Mallika et al. (2020) ‘Performance analysis of Neural Networks with IRI-2016 and IRI-2012 models over Indian low-latitude GPS stations’ https://link.springer.com/article/10.1007/s10509-020-03821-6

[4] Sivavaraprasad et al. (2020) ‘Performance evaluation of neural network TEC forecasting models over equatorial low-latitude Indian GNSS station. https://www.sciencedirect.com/science/article/pii/S1674984719300242

[5] Sivavaraprasad G, VenkataRatnam D, Sridhar M, Sivakrishna K ‘Modelling and forecasting of ionospheric TEC irregularities over a low latitude GNSS station’ https://link.springer.com/article/10.1007/s10509-020-03883 6

[6] Suneetha Emmelaa, Rama Laharia V, Anushaa B, Bhavanaa D, Yury V. Yasyukevich, Vladislav, Demyanov V, Venkata Ratnam D (2024) ‘Global ionospheric total electron content short-term forecast based on Light Gradient Boosting Machine, Extreme Gradient Boosting, and Gradient Boost Regression’ https://www.sciencedirect.com/science/article/abs/pii/S0273117724007063

[7] Veera Kumar Maheswarana, James A. Baskaradasa, Raju Nagarajana Rajesh Anbazhagana, Sriram Subramaniana, Venkata Ratnam Devanaboyinab, Rupesh M. Das ‘Bi-LSTM based vertical total electron content prediction low-latitude equatorial ionization anomaly region of South India.
https://www.sciencedirect.com/science/article/abs/pii/S0273117723007020

Indra Ganesan

★ COLLEGE OF ENGINEERING ★

(AN AUTONOMOUS INSTITUTION)

Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
NAAC Accredited, 2 (f) & 12 (B) Status Institution by UGC
IG Valley (Near New Integrated Bus Terminal), Manikandam, Tiruchirappalli - 620 012

IQAC RDDB

7th ANRF Sponsored International Conference on

Artificial Intelligence Data Science and Cyber Security

(HYBRID MODE)

Certificate of Participation

This is to certify that Mr. / Ms. Ambika K / SCE

Harnessing AI-Driven Predictions of Total Electron Content in Ionosphere

has actively participated/presented in ANRF Sponsored 7th International Conference on **"Artificial Intelligence, Data Science and Cyber Security"** held on 3rd & 4th April 2025.

Dr. S. Karthikeyan	Dr. M. Anuradha	Dr. G. Balakrishnan	Dr. G. Rajasekaran
Advisory Committee	Advisory Committee	Convenor	Patron


Indra Ganesan
★ COLLEGE OF ENGINEERING ★
(AN AUTONOMOUS INSTITUTION)
Hi Valley (Near New Integrated Bus Terminal), Manikandam, Tiruchirappalli - 620 012
IQAC
RDDB
7th
ANBF Sponsored
International Conference on
Artificial Intelligence Data Science and Cyber Security
(HYBRID MODE)
Certificate of Participation
This is to certify that Mr / Ms. Janani D , SCE
Harnessing AI-Driven Predictions of Total Electron Content in Ionosphere
has actively participated/presented in ANBF Sponsored 7th International Conference on "Artificial Intelligence, Data Science and Cyber Security" held on 3rd & 4th April 2025.
Dr. S. Karthikeyan
Advisory Committee
Dr. M. Anuroja
Advisory Committee
Dr. G. Balakrishnan
Convenor
Er. G. Rajasekaran
Patron


Indra Ganesan
★ COLLEGE OF ENGINEERING ★
(AN AUTONOMOUS INSTITUTION)
Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
NAAC Accredited, 2 (f) & 12 (B) Status Institution by UGC
ID Valley (Near New Integrated Bus Terminal), Manikandam, Tiruchirappalli - 620 012
IQAC
RDDB
7th
ANRF Sponsored
International Conference on
Artificial Intelligence Data Science and Cyber Security
(HYBRID MODE)
Certificate of Participation
This is to certify that Mr. / Ms. Rupa Sri Varsini R / SCE
Harnessing AI-Driven Predictions of Total Electron Content in Ionosphere
has actively participated/presented in ANRF Sponsored 7th International Conference on "Artificial Intelligence, Data Science and Cyber Security" held on 3rd & 4th April 2025.
Dr. S. Karthikeyan
Advisory Committee
Dr. M. Arunraja
Advisory Committee
Dr. G. Balakrishnan
Convenor
Er. G. Rajasekaran
Patron


Indra Ganesan
★ COLLEGE OF ENGINEERING ★
(AN AUTONOMOUS INSTITUTION)
Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
NBA Accredited, 2 (f) & 12 (B) Status Institution by UGC
IQ Valley (Near New Integrated Bus Terminal), Manikandam, Tiruchirappalli - 620 012
IQAC
RDDB
7th
ANRF Sponsored
International Conference on
Artificial Intelligence Data Science and Cyber Security
(HYBRID MODE)
Certificate of Participation
This is to certify that Mr. / Ms. Sridevi A / SCE
Harnessing AI-Driven Predictions of Total Electron Content in Ionosphere
has actively participated/presented in ANRF Sponsored 7th International Conference on "Artificial Intelligence, Data Science and Cyber Security" held on 3rd & 4th April 2025
Dr. S. Karthikeyan
Advisory Committee
Dr. M. Arunraja
Advisory Committee
Dr. G. Balakrishnan
Convenor
Er. G. Rajaselvaraj
Patron