Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Machine Learning-Based Prediction of LTE Uplink Interference using Atmospheric Ducting and Weather Features

Imran Kamal Mohammed Kadri
Radio Frequency Optimization and Data Analytics
Eastern Province, Saudi Arabia

***Abstract*— Uplink interference remains a major challenge in LTE radio access networks because of its highly dynamic temporal and spatial behavior. In TDD systems, atmospheric ducting can extend the propagation range of downlink transmissions from distant base stations beyond normal limits, causing them to overlap with the uplink reception period of neighboring cells resulting in severe uplink interference. This paper investigates whether meteorological conditions associated with atmospheric ducting can be used to predict uplink interference in a commercial LTE network in the Eastern Province of Saudi Arabia. Historical weather data, including pressure-level atmospheric variables, were integrated with one year of operational support system (OSS) measurements from more than 1,100 LTE sites. Physics-informed features derived from temperature, humidity, refractivity, atmospheric stability, and wind profiles were combined with network performance indicators to train multiple machine learning models. Experimental results show that conventional OSS indicators alone provide limited predictive capability ($R^2 \approx 0.03$), whereas atmospheric variables substantially improve prediction performance. The best results were obtained by combining weather and OSS features, with a Random Forest model achieving an R^2 of approximately 0.75. Feature importance analysis identified temperature, temperature inversion, refractivity gradients, and pressure-level atmospheric characteristics as key predictors, supporting the hypothesis that atmospheric conditions associated with tropospheric ducting are strongly associated with uplink interference and provide substantial predictive value. The proposed framework demonstrates a practical approach for integrating numerical weather data with cellular network measurements, enabling proactive interference prediction and providing a foundation for weather-aware radio network optimization.**

Keywords—*LTE, uplink interference, tropospheric ducting, atmospheric refractivity, machine learning, Random Forest, ERA5, Open-Meteo, radio access networks, weather analytics, RF optimization.*

I. NTRODUCTION

The complexity of radio network optimization has significantly increased after the introduction of Long-Term Evolution (LTE) mobile communication systems. Modern cellular networks employ advanced technologies including carrier aggregation, massive multiple-input multiple-output (Massive MIMO), beamforming, and dynamic spectrum utilization to maximize spectral efficiency and user experience [1], [3], [4]. While these technologies have substantially improved network capacity, maintaining stable radio performance under varying environmental conditions remains a major challenge for network operators.

Among the various radio impairments encountered in operational networks, uplink (UL) interference represents one of the most difficult problems to diagnose and mitigate. Elevated uplink interference reduces receiver sensitivity at the base station, resulting in lower uplink throughput, reduced signal quality, increased retransmissions, and degradation of customer experience. Conventional optimization approaches primarily attribute uplink interference to factors such as excessive traffic loading, external radio sources, hardware faults, or network configuration issues. However, experienced radio optimization engineers frequently observe periods during which severe uplink interference occurs simultaneously across geographically separated sites without any corresponding changes in network configuration or traffic behavior. Such observations suggest that additional environmental mechanisms may influence uplink interference beyond traditional radio network parameters.

One atmospheric phenomenon that has long attracted attention within the radio propagation community is tropospheric ducting. Under specific meteorological conditions, abnormal vertical gradients of temperature and moisture alter the refractive index of the lower atmosphere, causing radio waves to bend toward the Earth's surface rather than propagating along their normal trajectories. This process effectively creates atmospheric ducts that can trap electromagnetic waves and allow them to travel hundreds of kilometers beyond normal line-of-sight distances [1], [17]. Although tropospheric ducting has been extensively studied for radar systems, microwave links, maritime communications, and amateur radio propagation [13], comparatively little research has investigated its impact on large-scale LTE cellular uplink interference using modern machine learning techniques.

Recent advances in publicly accessible numerical weather reanalysis datasets and machine learning algorithms provide an opportunity to bridge this gap. Global atmospheric datasets such as ERA5 and Open-Meteo now enable retrieval of both surface meteorological observations and pressure-level atmospheric profiles with sufficient temporal and spatial resolution to investigate their relationship with

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

cellular network performance [9], [10], [14], [15]. Simultaneously, ensemble learning algorithms such as Random Forest and Extra Trees have demonstrated strong capability for modelling highly nonlinear relationships within heterogeneous engineering datasets [6], [7].

Despite these developments, several practical challenges remain unresolved. Atmospheric datasets are generally available on coarse spatial grids, whereas cellular networks consist of thousands of geographically distributed sites. Furthermore, meteorological observations are available at multiple temporal resolutions, while network performance indicators may be recorded hourly or daily. The integration of these heterogeneous datasets requires careful spatial matching, temporal synchronization, and feature engineering before meaningful predictive modelling can be performed.

The present study originated from repeated operational observations within a commercial LTE network in the Eastern Province of Saudi Arabia, where uplink interference frequently exhibited temporal patterns that closely resembled known atmospheric ducting conditions observed in coastal environments. Rather than assuming a direct causal relationship, the research adopted a systematic hypothesis-driven methodology aimed at determining whether atmospheric variables could explain or predict variations in uplink interference.

Unlike many machine learning studies that present only successful experiments, this paper documents the complete evolution of the investigation. Multiple data acquisition strategies, feature engineering approaches, predictive models, and validation methodologies were evaluated before arriving at the final framework. Several promising approaches proved ineffective, necessitating substantial redesign of the methodology. Documenting these unsuccessful experiments is considered an important contribution of this work because they illustrate the practical challenges associated with integrating atmospheric science and radio network optimization.

Ultimately, this study demonstrates that atmospheric variables — particularly those associated with vertical refractivity structure — provide valuable predictive information beyond conventional OSS performance indicators. The findings suggest that incorporating meteorological intelligence into future self-organizing network (SON) platforms may enable more proactive identification and mitigation of uplink interference events, particularly within coastal deployments susceptible to abnormal radio propagation.

Fig. 1 illustrates this mechanism schematically before the underlying propagation physics and prior literature are reviewed in detail.

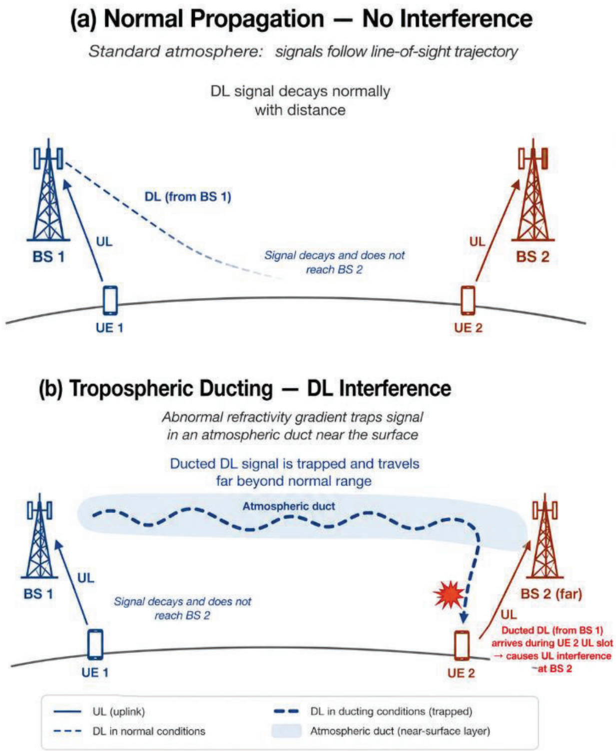

Fig. 1. Mechanism of tropospheric-ducting-induced uplink interference. Under normal propagation (a), downlink signals decay with distance and do not reach distant uplink receivers. Under ducting conditions (b), an abnormal refractivity gradient traps the downlink signal near the surface, allowing it to travel far beyond its normal range and arrive at a distant base station during a neighbouring uplink slot, producing uplink interference.

II. LITERATURE REVIEW

The increasing complexity of Long-Term Evolution (LTE) radio access networks has significantly expanded the scope of radio network optimization beyond traditional coverage and capacity planning. Modern cellular systems operate under highly dynamic propagation environments in which user mobility, traffic demand, antenna beamforming, and atmospheric conditions jointly influence radio performance. Consequently, researchers have increasingly explored the application of data-driven approaches for predicting key network performance indicators (KPIs), including throughput, latency, handover success rate, call drops, and interference. Nevertheless, comparatively little attention has been devoted to understanding the influence of atmospheric propagation phenomena on uplink interference.

A. Uplink Interference in TDD LTE Networks

Unlike Frequency Division Duplex (FDD) systems, where uplink and downlink transmissions operate on separate frequency bands, Time Division Duplex (TDD) systems utilize the same carrier frequency for both uplink (UL) and downlink (DL) transmissions, separating them only in the time domain through synchronized UL/DL slot configurations [3], [18]. Under normal propagation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

conditions, neighboring cells follow identical transmission timing, ensuring that downlink transmissions from one cell do not overlap with uplink receptions in another.

However, abnormal propagation conditions can disrupt this assumption. During tropospheric ducting events, downlink signals transmitted by distant base stations may propagate hundreds of kilometers beyond their normal coverage area. Because of the extended propagation delay, these high-power downlink signals can arrive during the uplink reception period of another TDD cell. As a result, the receiving base station interprets the delayed downlink transmission as uplink interference.

This form of interference differs fundamentally from conventional uplink interference generated by user equipments (UEs). UE transmissions are power-controlled and typically limited to a maximum transmit power of approximately 23 dBm, whereas base stations transmit downlink signals at powers exceeding 40–46 dBm [3], [18]. Consequently, delayed downlink transmissions from distant cells can produce significantly stronger interference than ordinary uplink signals, substantially degrading uplink reception.

Since these interference events originate from atmospheric propagation rather than changes in traffic load or network configuration, they are often not explained by conventional Operational Support System (OSS) performance indicators, making prediction and mitigation particularly challenging.

Several studies have shown that excessive uplink interference results in:

- Reduced uplink SINR
- Increased block error rate
- Lower uplink throughput
- Higher HARQ retransmissions
- Degradation of cell-edge performance
- Reduced overall network capacity

Conventional literature attributes uplink interference primarily to network-related causes, including excessive traffic loading, improper uplink power control, hardware failures, external radio frequency sources, neighboring cell interference, and scheduling inefficiencies. These factors are typically observable through Operational Support System (OSS) performance counters and therefore constitute the basis of most commercial optimization strategies.

However, field engineers frequently report situations in which elevated uplink interference simultaneously affects large geographical regions despite relatively stable network traffic and unchanged radio configurations. Such observations indicate that additional environmental mechanisms may influence uplink interference but are not represented within conventional OSS statistics.

B. Tropospheric Ducting and Radio Wave Propagation

Radio wave propagation within the lower atmosphere is governed by spatial variations in atmospheric refractivity [11]. Under standard atmospheric conditions, radio waves gradually bend toward the Earth's surface due to the normal decrease in refractive index with altitude. Certain meteorological conditions, however, produce abnormal refractivity gradients capable of trapping electromagnetic waves within narrow atmospheric layers. This phenomenon, commonly known as tropospheric ducting, allows radio signals to propagate well beyond the normal radio horizon.

Tropospheric ducts typically develop under conditions involving:

- Strong temperature inversions
- Rapid decreases in atmospheric moisture with altitude
- Stable high-pressure systems
- Weak vertical mixing
- Coastal sea–land temperature contrasts
- Low-level marine boundary layers

These conditions are particularly common in coastal regions surrounding the Arabian Gulf, where intense solar heating combined with marine air masses frequently produces highly stable atmospheric layers during summer months.

Previous propagation studies have demonstrated that atmospheric ducting significantly increases transmission distance for radar systems, microwave backhaul, maritime communications, television broadcasting, VHF/UHF radio systems, and satellite communication links [1], [13], [17]. Although the underlying physics is well established, quantitative studies examining ducting effects within operational LTE cellular networks remain relatively limited.

C. Atmospheric Refractivity as a Propagation Indicator

Atmospheric refractivity is commonly expressed through the radio refractivity parameter N, defined as a function of atmospheric pressure, temperature, and water vapour pressure [11]. To account for Earth curvature, propagation studies frequently employ modified refractivity (M) rather than refractivity alone [12]. The vertical gradients dN/dh and dM/dh are widely recognized as important indicators of abnormal propagation. Strong negative modified refractivity gradients indicate atmospheric conditions favourable for trapping radio waves within surface or elevated ducts.

Previous studies have demonstrated that refractivity profiles derived from radiosonde observations or numerical weather prediction models provide considerably more reliable indicators of duct formation than surface weather observations alone [1], [13]. This distinction is particularly important because atmospheric ducting is fundamentally a three-dimensional phenomenon. Surface temperature or humidity measurements alone cannot fully describe the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

vertical atmospheric structure responsible for anomalous propagation.

D. Numerical Weather Reanalysis for Telecommunications Research

The availability of high-resolution atmospheric reanalysis datasets has significantly expanded opportunities for interdisciplinary research combining meteorology and wireless communications. Among the most widely used datasets is the ERA5 reanalysis, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) [9], [10]. ERA5 provides hourly global atmospheric estimates by assimilating observations from satellites, aircraft, weather stations, ships, and radiosondes into numerical weather prediction models [10].

ERA5 offers several advantages for telecommunications research, including global spatial coverage, hourly temporal resolution, multiple atmospheric pressure levels, physically consistent meteorological fields, and long historical records. Pressure-level variables available within ERA5 enable computation of refractivity, modified refractivity, temperature inversion, moisture gradients, wind shear, and atmospheric stability — variables directly related to radio propagation mechanisms and therefore substantially richer than conventional surface weather observations.

More recently, publicly accessible services such as Open-Meteo have simplified retrieval of both historical weather observations and archived numerical weather forecasts through REST-based application programming interfaces (APIs) [14], [15]. These services eliminate many of the practical barriers associated with downloading large reanalysis datasets while maintaining sufficient spatial and temporal resolution for engineering applications.

E. Machine Learning in Cellular Network Optimization

Machine learning has become an increasingly important tool for mobile network optimization due to its ability to model highly nonlinear relationships among heterogeneous variables [5]. Common applications include traffic prediction, throughput estimation, mobility optimization, anomaly detection, self-organizing networks (SON), energy optimization, customer experience prediction, and fault diagnosis.

Among supervised learning algorithms, ensemble tree methods have demonstrated particularly strong performance for radio network applications because they naturally capture nonlinear interactions while remaining relatively robust to multicollinearity and missing data. Random Forest has become one of the most widely adopted algorithms due to its high predictive accuracy, resistance to overfitting, computational efficiency, and straightforward feature importance estimation [6]. Several studies have also reported strong performance using Extra Trees [7], Decision Trees, Gradient Boosting [8], XGBoost, and LightGBM.

Despite these advances, most existing studies rely almost exclusively on network-generated counters collected from OSS platforms. Environmental variables are rarely incorporated, limiting the ability of these models to capture propagation phenomena driven by atmospheric conditions.

F. Weather-Based Prediction of Cellular Performance

A limited number of recent investigations have explored the influence of weather on cellular network performance [1], [17]. Reported meteorological variables include air temperature, humidity, rainfall, atmospheric pressure, wind speed, and solar radiation. These studies generally demonstrate that adverse weather conditions can influence received signal strength, propagation loss, throughput, latency, and handover performance.

However, most published work employs only surface weather observations, which provide limited information regarding the vertical atmospheric structure responsible for abnormal propagation. Furthermore, few studies attempt to derive physically meaningful propagation indicators such as refractivity gradients or temperature inversion from atmospheric pressure-level data. Consequently, many weather-based prediction models remain empirical rather than physics-informed.

III. RESEARCH GAP

Although extensive literature exists independently on atmospheric propagation and machine learning for cellular network optimization, several important research gaps remain.

First, most machine learning models developed for LTE optimization rely exclusively on OSS-derived performance counters. Environmental variables associated with atmospheric propagation are generally omitted despite longstanding evidence that abnormal propagation can significantly influence radio performance.

Second, previous weather-based investigations predominantly employ surface meteorological observations. While these variables describe local weather conditions, they do not capture the vertical refractivity gradients responsible for tropospheric duct formation.

Third, very few studies combine atmospheric physics with operational cellular network data collected from large commercial deployments. Existing work is often based on simulations, drive-test campaigns, or limited geographical case studies rather than long-term operational measurements.

Fourth, published studies rarely document the complete experimental process. Unsuccessful feature engineering strategies, alternative datasets, and methodological redesigns are typically omitted, making it difficult for future researchers to reproduce or extend the work.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Finally, to the authors' knowledge, no published study has systematically evaluated the predictive contribution of conventional OSS counters, surface weather variables, vertical atmospheric indicators, and spatial information within a unified machine learning framework for uplink interference prediction in a commercial LTE network.

Accordingly, this research seeks to bridge these gaps by integrating operational OSS statistics with atmospheric observations and ducting-related meteorological features derived from numerical weather datasets while documenting the complete evolution of the methodology from initial hypothesis to final predictive framework.

IV. RESEARCH OBJECTIVES

The primary objective of this study is to investigate whether atmospheric conditions associated with tropospheric ducting can improve the prediction of LTE uplink interference beyond conventional OSS-based approaches. The specific objectives are:

1. To investigate the relationship between atmospheric conditions and uplink interference in a commercial LTE network.
2. To construct a unified dataset by integrating OSS performance counters with weather observations and pressure-level atmospheric variables.
3. To derive ducting-related meteorological features, including refractivity gradients, modified refractivity, temperature inversion, humidity gradients, and wind shear.
4. To quantify the individual predictive contributions of location, OSS statistics, surface weather, and vertical atmospheric features.
5. To establish a practical workflow that can be integrated into future intelligent radio network optimization platforms.

V. RESEARCH JOURNEY AND METHODOLOGY EVOLUTION

Unlike conventional machine learning studies that present only the final predictive model, this research adopted an iterative hypothesis-driven methodology. The investigation evolved through multiple stages over several months, with each experiment informing subsequent methodological decisions. Several promising approaches proved unsuccessful or only partially successful, necessitating substantial redesign of the data acquisition strategy, feature engineering process, and modelling framework.

The chronological evolution of the research is illustrated in Fig. 2, while Table I summarizes the major experimental milestones.

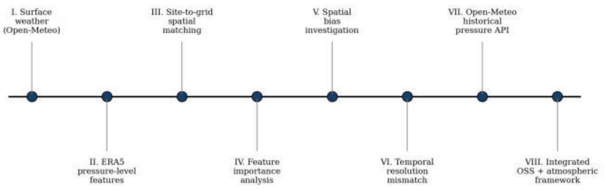

Fig. 2. Chronological evolution of the research methodology, from initial hypothesis formulation through Open-Meteo surface weather, ERA5 pressure-level integration, site-level weather matching, feature engineering, machine learning experiments, de-biasing, and the final predictive framework.

A. Initial Research Hypothesis

The research originated from repeated operational observations within a commercial LTE network deployed in the Eastern Province of Saudi Arabia. During routine radio optimization activities, periods of elevated uplink interference were frequently observed across geographically separated LTE cells despite the absence of significant traffic fluctuations, hardware alarms, or configuration changes. These observations suggested that conventional OSS statistics alone could not fully explain the temporal behaviour of uplink interference.

Since the affected region is located along the Arabian Gulf coastline, atmospheric ducting was considered a plausible contributing mechanism. Coastal environments frequently experience stable marine boundary layers and strong temperature inversions, both of which favour abnormal radio propagation. Accordingly, the initial research hypothesis was formulated as follows:

> *H_1: Atmospheric conditions associated with tropospheric ducting contribute significantly to variations in LTE uplink interference and can be exploited to improve predictive machine learning models.*

Rather than attempting to prove causality directly, the research focused on determining whether meteorological variables could improve the prediction of uplink interference beyond conventional OSS measurements.

B. Phase I — Surface Weather-Based Prediction

The first phase of the investigation employed publicly available weather observations obtained through the Open-Meteo Archive API. The motivation for selecting Open-Meteo was threefold: (1) free public access without authentication, (2) historical weather availability, and (3) a straightforward REST API suitable for large-scale automation.

Initially, weather data were collected for a representative location within the study area. The downloaded variables included air temperature, relative humidity, dew point, atmospheric pressure, wind speed, and wind gusts. These variables were synchronized with hourly uplink interference measurements exported from the network OSS platform.

Several meteorological indicators were derived from the downloaded weather observations, including daily temperature range, dew-point spread, pressure variation, temporal encoding using hour-of-day, and cyclic sine/cosine transformations. The expectation was that these variables would indirectly capture atmospheric stability associated with duct formation.

Multiple regression algorithms were evaluated, including Linear Regression, Decision Tree, Random Forest, Extra Trees, and Gradient Boosting. Among these models, Random Forest consistently achieved the highest predictive performance; however, the overall prediction accuracy remained modest. Feature importance analysis revealed that surface weather variables contributed to the prediction, but substantial unexplained variability remained.

This finding led to the first major realization of the study: surface weather observations describe atmospheric conditions near the ground but do not adequately characterize the vertical atmospheric structure responsible for tropospheric ducting. Consequently, the research direction shifted toward incorporating atmospheric pressure-level information.

C. Phase II — Incorporating Atmospheric Physics Using ERA5

To better represent the physical mechanisms responsible for abnormal radio propagation, the second phase introduced ERA5 atmospheric reanalysis data. Unlike surface weather observations, ERA5 provides meteorological variables at multiple atmospheric pressure levels, enabling calculation of propagation-related quantities used in radio propagation research.

The downloaded pressure-level data were transformed into several physically meaningful propagation indicators, including atmospheric refractivity (N), modified refractivity (M), refractivity gradient (dN/dh), modified refractivity gradient (dM/dh), temperature inversion, relative humidity gradient, and wind shear. Unlike surface weather variables, these quantities directly characterize the vertical atmospheric structure responsible for anomalous propagation. The expectation was that these features would provide stronger predictive capability than conventional weather observations alone.

D. Site-Level Weather Assignment

Because ERA5 is distributed on a regular geographical grid rather than at cellular site locations, a nearest-grid spatial matching procedure was developed to assign atmospheric observations to each of the more than 1,100 LTE sites and merge them with OSS measurements by timestamp and site identifier (full procedure detailed in Section VI.D). This produced a unified dataset linking atmospheric observations with operational network performance across the entire study area.

E. Initial Success and an Unexpected Problem

Introducing pressure-level atmospheric variables produced a noticeable improvement in predictive performance. However, feature importance analysis revealed an unexpected result: latitude and longitude became the dominant predictors. Instead of learning atmospheric behaviour, the model was primarily learning site identity. The highest feature importances were consistently assigned to latitude and longitude, while atmospheric variables received substantially lower importance.

This finding indicated that certain sites naturally experienced higher baseline interference than others. Although spatial information improved predictive accuracy, it obscured the contribution of atmospheric variables.

F. Investigating Spatial Bias

To better understand the influence of geographical information, several additional experiments were performed. Latitude and longitude were removed from the feature set, after which model performance decreased substantially. This observation confirmed that spatial information captured significant variability within the interference measurements.

Rather than concluding that weather was unimportant, the results suggested that interference depends on two distinct components: (1) a relatively stable spatial baseline associated with individual sites, and (2) temporal fluctuations influenced by atmospheric conditions. This distinction guided subsequent feature engineering efforts.

G. Temporal Resolution Issue

While analysing the prediction results, another important limitation became apparent: The atmospheric pressure-level dataset contained approximately one month of hourly observations. This created two significant challenges. First, the atmospheric dataset did not capture seasonal variability. Second, machine learning models trained on one month of data exhibited poor generalization when evaluated over longer periods. Chronological validation demonstrated that although random train-test splits produced encouraging results, predictive accuracy deteriorated when future time periods were evaluated. The investigation concluded that the limited temporal diversity of the atmospheric data prevented robust long-term learning.

H. Alternative Data Acquisition Strategies

Several approaches were investigated to overcome the limited availability of pressure-level observations. The first solution considered downloading extended ERA5 pressure-level datasets covering an entire year. Although technically feasible, this approach proved impractical due to very large download sizes, long processing times, storage requirements, and repeated API limitations encountered during retrieval.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

Consequently, an alternative strategy was explored. The Open-Meteo Historical Forecast API was found to provide archived pressure-level forecasts for multiple atmospheric levels through a significantly simpler interface. Unlike the original Open-Meteo archive dataset, which primarily provides surface observations, the historical forecast service exposed pressure-level temperature, humidity, dew point, geopotential height, and wind speed, allowing the same refractivity calculations previously performed using ERA5. This discovery substantially simplified the overall data acquisition workflow while preserving the required atmospheric information.

I. Final Research Direction

The final methodology therefore combined the advantages of both approaches. Surface weather observations and pressure-level atmospheric profiles were retrieved using automated API calls, spatially matched to LTE sites, transformed into physically meaningful atmospheric indicators, and integrated with OSS measurements for machine learning.

The resulting dataset simultaneously represented conventional network performance, geographical characteristics, surface meteorology, and vertical atmospheric structure. This evolution — from simple weather observations to a physics-informed atmospheric framework — formed the foundation of the final predictive model evaluated in the remainder of this paper.

Table I. Chronological Summary of Research Evolution

Phase	Objective	Outcome	Decision
I	Surface weather from Open-Meteo	Moderate prediction accuracy; insufficient atmospheric representation	Introduce pressure-level variables
II	ERA5 pressure-level features	Improved physical realism	Develop refractivity-based indicators
III	Site-to-grid spatial matching	Enabled weather assignment for >1,100 LTE sites	Build integrated dataset
IV	Feature importance analysis	Latitude/longitude dominated predictions	Investigate spatial bias
V	Remove spatial features	Significant accuracy reduction	Retain spatial context but analyze contributions separately
VI	One-month hourly modelling	Poor temporal generalization	Seek longer historical atmospheric data
VII	Full-year pressure-level retrieval	ERA5 downloads impractical	Transition to Open-Meteo historical pressure API
VIII	Integrated atmospheric and OSS framework	Final prediction pipeline established	Proceed to comprehensive ML evaluation

VI. DATASET DESCRIPTION AND FEATURE ENGINEERING

The effectiveness of any machine learning model depends fundamentally on the quality and representativeness of the input data. In this study, a unified dataset was constructed by integrating operational cellular network measurements with atmospheric observations and engineered propagation indicators. Since the required information originated from multiple heterogeneous sources with different spatial and temporal resolutions, considerable effort was devoted to data acquisition, synchronization, preprocessing, and feature engineering.

The final dataset comprised four principal categories of information: (1) Operational Support System (OSS) network statistics, (2) surface meteorological observations, (3) pressure-level atmospheric variables, and (4) engineered ducting-related and temporal features. The overall data preparation workflow is illustrated in Fig. 3.

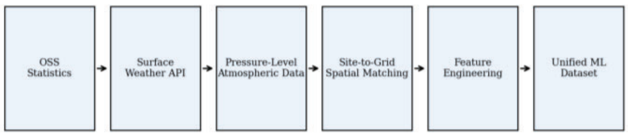

Fig. 3. Overall data acquisition and preprocessing pipeline: OSS statistics, surface weather API, and pressure-level atmospheric data are spatially matched to LTE sites, engineered into propagation features, and merged into the unified ML dataset.

A. Operational Support System (OSS) Dataset

The primary source of network performance information was the commercial Operational Support System (OSS) used for LTE network optimization. The OSS records hourly and daily performance indicators for every operational site within the study area.

The target variable selected for this research was uplink interference, expressed using the operator's standard interference measurement. This metric reflects the average interference level received by the base station and is routinely monitored by radio optimization engineers to identify cells experiencing degraded uplink performance.

In addition to the target variable, several OSS counters describing network utilization were incorporated as

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

explanatory variables. Following feature selection and correlation analysis, the final OSS variables included uplink Physical Resource Block (UL PRB) utilization and uplink traffic (L.Thrp.bits.UL). Initially, additional traffic-related variables were considered, including downlink throughput, average user traffic, and downlink PRB utilization; however, correlation analysis revealed substantial redundancy among these features, leading to their removal during later stages of the study.

Table II. Operational Support System Variables

Variable	Description	Unit
UL Interference	Target variable	dBm
UL PRB	Uplink Physical Resource Block utilization	%
L.Thrp.bits.UL	Uplink traffic	bit/s

B. Surface Meteorological Dataset

Historical meteorological observations were initially obtained using the Open-Meteo Archive API. The following variables were retrieved: air temperature (2 m), relative humidity (2 m), dew point temperature (2 m), mean sea-level pressure, wind speed (10 m), and wind gust speed (10 m).

These variables were selected because previous propagation studies have associated them with atmospheric stability and radio propagation characteristics.

C. Pressure-Level Atmospheric Dataset

To characterize the vertical structure of the lower atmosphere, pressure-level meteorological variables were incorporated into the dataset. Unlike surface weather observations, pressure-level variables describe atmospheric conditions at multiple altitudes, allowing estimation of refractivity gradients and atmospheric stability.

The following pressure levels were included: 1000, 975, 950, 925, 900, and 850 hPa. For each pressure level, temperature, relative humidity, dew point temperature, wind speed, and geopotential height were extracted. These observations enabled computation of atmospheric quantities directly related to radio propagation.

D. Site-Level Weather Assignment

Weather datasets are provided on regularly spaced geographical grids, whereas cellular network sites are located at irregular positions. Consequently, atmospheric observations could not be used directly, and a spatial matching procedure was developed. For every LTE site: (1) site latitude and longitude were obtained from the network database, (2) the nearest atmospheric grid point was identified, (3) weather observations from that grid point were assigned to the site, and (4) weather records were synchronized with OSS timestamps.

This process was repeated for more than 1,100 operational LTE sites, enabling construction of a site-specific atmospheric dataset. Although multiple neighbouring sites occasionally shared the same weather grid, geographical coordinates remained unique, preserving spatial differentiation across the network.

E. Data Cleaning and Preprocessing

Prior to model training, several preprocessing operations were applied to improve data quality and ensure consistency across all variables. The preprocessing workflow included removal of duplicate records, elimination of incomplete observations, timestamp standardization, spatial matching validation, numerical type conversion, and missing value imputation where appropriate.

Outliers were intentionally retained unless identified as clear measurement errors because extreme atmospheric conditions were considered particularly relevant for uplink interference prediction. Continuous variables were inspected for unrealistic values resulting from API errors or missing atmospheric observations.

The final dataset contained 1,128,841 observations collected from 1,100 LTE sites over one year.

F. Feature Engineering

Feature engineering constituted one of the most important stages of the study. Rather than relying solely on directly observed weather variables, several derived features were introduced to better represent atmospheric processes associated with tropospheric ducting. The engineered variables can be categorized into three groups.

1) Surface Weather Features: Daily statistical summaries were generated from hourly weather observations, including mean, maximum, and minimum temperature, mean humidity, mean dew point, mean pressure, mean wind speed, and maximum wind gust. Additional derived variables included daily temperature range, pressure change, and dew-point spread. These features were intended to represent daily atmospheric variability.

2) Atmospheric Propagation Features: The pressure-level observations enabled calculation of several physically meaningful propagation indicators, including atmospheric refractivity (N), computed from pressure, temperature, and water vapour pressure; modified refractivity (M), incorporating Earth curvature effects; the refractivity gradient (dN/dh) and modified refractivity gradient (dM/dh), representing the vertical rate of change of refractivity and one of the principal indicators of duct formation; temperature inversion, defined as the difference between temperatures at adjacent pressure levels; relative humidity gradient, representing vertical moisture variation; and wind shear, the difference in wind speed between adjacent atmospheric layers. For each quantity, both mean and extreme values were evaluated during feature engineering.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

3) Ducting Indicators: To summarize prolonged abnormal atmospheric conditions, additional ducting-related features were created, including the number of ducting hours, the daily ducting ratio, and the maximum and mean inversion strength. These variables were designed to capture the persistence of favourable ducting conditions rather than isolated hourly events.

Table III. Engineered Atmospheric Features

Category	Features
Surface Weather	Temperature, Humidity, Dew Point, Pressure, Wind Speed, Wind Gust
Derived Weather	Temperature Range, Dew Spread, Pressure Change
Propagation	N, M, dN/dh, dM/dh
Stability	Temperature Inversion, RH Gradient, Wind Shear
Ducting	Ducting Hours, Ducting Ratio

G. Correlation Analysis and Feature Selection

Before model development, Pearson correlation analysis was performed to identify redundant predictors. Several variables exhibited extremely high correlation coefficients, indicating that they contributed similar information to the model. Examples included maximum and minimum temperature, ducting hours and ducting ratio, mean and maximum inversion strength, and multiple traffic-related OSS counters.

Removing highly correlated variables reduced model complexity while preserving predictive performance. Interestingly, Random Forest models exhibited only a marginal reduction in prediction accuracy after redundant features were removed, confirming that the simplified feature set retained the essential information. The final dataset therefore represented a balance between predictive capability, interpretability, and computational efficiency.

H. Final Dataset Composition

Following preprocessing and feature engineering, each observation within the final dataset consisted of a site identifier, geographic coordinates, OSS utilization metrics, surface weather observations, pressure-level atmospheric variables, engineered propagation indicators, temporal descriptors, and the uplink interference measurement. This integrated dataset formed the basis for the machine learning experiments described in the following section.

VII. MACHINE LEARNING METHODOLOGY

Following the integration of OSS statistics, meteorological observations, and engineered atmospheric features, a supervised machine learning framework was developed to predict uplink interference. The objective was not merely to identify the algorithm with the highest prediction accuracy, but also to understand the relative contribution of different feature groups and evaluate whether atmospheric information provides predictive value beyond conventional network performance indicators.

To achieve this objective, multiple regression algorithms were systematically evaluated under identical experimental conditions, allowing direct comparison of their predictive capability, robustness, and interpretability.

A. Overall Modelling Framework

The complete machine learning workflow adopted in this study is illustrated in Fig. 4. The modelling process consisted of data preprocessing and cleaning, feature engineering, correlation-based feature reduction, train-test dataset partitioning, model training, hyperparameter optimization, performance evaluation, feature importance analysis, and comparative assessment of feature groups.

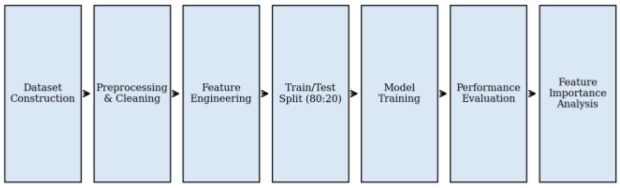

Fig. 4. Machine learning workflow for uplink interference prediction, from dataset construction through preprocessing, feature engineering, model training, evaluation, and feature importance analysis.

B. Target Variable

The supervised learning task was formulated as a regression problem. The target variable was the measured uplink interference obtained from the commercial OSS platform. Unlike classification-based approaches that simply categorize interference into discrete severity levels, regression preserves the continuous nature of interference measurements and provides greater flexibility for practical network optimization. Predicting continuous interference values enables network engineers to estimate future interference magnitude and prioritize optimization actions before customer experience is significantly degraded.

C. Feature Groups

To better understand the influence of different information sources, several independent feature groups were evaluated throughout the research: (1) OSS features such as UL PRB utilization and uplink traffic; (2) weather features such as temperature, humidity, pressure, wind speed, wind gust, and dew point; (3) atmospheric propagation features such as dN/dh, dM/dh, temperature inversion, wind shear, humidity gradient, and ducting ratio; (4) spatial features, namely latitude and longitude; and (5) the combined feature set incorporating all available categories. Evaluating these groups independently enabled quantitative assessment of the predictive contribution of atmospheric information relative to conventional OSS statistics.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

D. Train-Test Strategy

The prepared dataset was divided into independent training and testing subsets. An 80:20 split was employed, whereby approximately 80% of the observations were used for model training and the remaining 20% were reserved for independent evaluation. The testing dataset remained completely unseen during model training. This strategy ensured that reported prediction accuracy reflected the model's ability to generalize rather than memorize training observations.

E. Machine Learning Algorithms

Five regression algorithms representing different modelling philosophies were evaluated.

1) Linear Regression: Served as the baseline model. Its simplicity makes it useful for determining whether uplink interference exhibits predominantly linear relationships with atmospheric variables. However, given the complex nonlinear interactions governing atmospheric propagation, Linear Regression was expected to provide limited predictive capability.

2) Decision Tree: Decision Trees recursively partition the feature space into regions exhibiting similar target values. Advantages include nonlinear modelling capability, interpretability, and automatic feature selection; however, individual decision trees are highly sensitive to noise and frequently overfit training data.

3) Random Forest: Random Forest constructs an ensemble of independently trained decision trees using bootstrap aggregation, with predictions obtained by averaging outputs across all trees [6]. Advantages include strong nonlinear modelling, reduced overfitting, robustness to noisy variables, resistance to multicollinearity, and feature importance estimation. Given these characteristics, Random Forest was expected to perform well for heterogeneous atmospheric datasets.

4) Extra Trees: Extra Trees (Extremely Randomized Trees) extend Random Forest by introducing additional randomization during tree construction [7]. Instead of searching exhaustively for the optimal split, candidate split points are selected randomly. This approach often improves computational efficiency while reducing variance, and Extra Trees have previously demonstrated competitive performance for environmental prediction problems involving nonlinear relationships.

5) Gradient Boosting: Gradient Boosting sequentially constructs weak learners that progressively reduce prediction error [8]. Unlike Random Forest, which trains trees independently, Gradient Boosting focuses on correcting residual errors generated by previous trees. Although Gradient Boosting often achieves high accuracy, it is generally more sensitive to hyperparameter selection and noisy data.

Table IV. Machine Learning Algorithms Evaluated

Model	Type	Characteristics
Linear Regression	Linear	Baseline model
Decision Tree	Tree-based	Simple nonlinear regression
Random Forest	Ensemble	Bagging, high robustness
Extra Trees	Ensemble	Randomized tree ensemble
Gradient Boosting	Ensemble	Sequential boosting

F. Hyperparameter Selection

Each algorithm was configured using commonly accepted hyperparameter settings suitable for regression tasks. Random Forest and Extra Trees were trained using multiple decision trees to improve prediction stability. Decision Tree depth was constrained to reduce overfitting. Gradient Boosting learning rate and tree depth were adjusted through iterative experimentation.

Although exhaustive hyperparameter optimization was beyond the scope of this study, multiple parameter combinations were evaluated to ensure fair comparison among models. The final configuration represented a balance between prediction accuracy, computational efficiency, and model interpretability.

Table V. Final Hyperparameter Configuration

Model	Important Hyperparameters
Linear Regression	Default
Decision Tree	max_depth = 15, min_samples_split = 3
Random Forest	n_estimators = 300, max_depth = 15
Extra Trees	n_estimators = 300
Gradient Boosting	learning_rate = 0.05, n_estimators = 300, max_depth = 3

G. Evaluation Metrics

Model performance was evaluated using three complementary regression metrics. The coefficient of determination (R^2) quantifies the proportion of variance explained by the predictive model:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

Higher values indicate better predictive performance. Since models with larger numbers of predictors may artificially inflate R^2, the adjusted coefficient of determination was also calculated; adjusted R^2 penalizes unnecessary predictors and therefore provides a more reliable comparison between feature sets of different dimensionalities. Finally, Mean Absolute Error (MAE) measures the average prediction error in the same units as the target variable:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i|$$

Unlike squared-error metrics, MAE is less sensitive to extreme observations and therefore provides an intuitive measure of prediction accuracy.

H. Feature Importance Analysis

Prediction accuracy alone does not explain why a machine learning model makes particular predictions. Consequently, feature importance analysis formed a central component of the investigation. For tree-based ensemble methods, feature importance was estimated using impurity reduction accumulated across all decision trees.

Feature importance analysis was used to identify dominant predictors, evaluate the contribution of atmospheric variables, investigate spatial bias, guide feature selection, and support physical interpretation of the predictive model. One of the most significant findings of the research emerged during this stage, where latitude and longitude consistently ranked above atmospheric variables. This observation motivated several subsequent experiments aimed at separating spatial effects from meteorological influences.

I. Comparative Experimental Design

To quantify the individual contribution of different information sources, four principal modelling scenarios were evaluated: Scenario 1 (OSS features only) determines whether conventional network counters alone can explain uplink interference; Scenario 2 (weather features only) evaluates the predictive capability of atmospheric observations without network information; Scenario 3 (location + weather) determines whether geographical context enhances atmospheric prediction; and Scenario 4 (location + weather + OSS) evaluates the complete integrated framework, representing the proposed methodology presented in this paper.

Table VI. Experimental Scenarios

Scenario	Feature Categories	Objective
S1	OSS only	Conventional baseline
S2	Weather only	Environmental prediction
S3	Weather + Location	Spatial environmental prediction
S4	Weather + Location + OSS	Final integrated framework

J. Model Interpretation Strategy

Rather than selecting the model with the highest R^2 alone, the evaluation considered four complementary criteria: prediction accuracy, model robustness, physical interpretability, and practical applicability for commercial radio optimization. This multi-dimensional evaluation ensured that the selected model not only achieved strong predictive performance but also provided insights into the atmospheric processes influencing uplink interference.

VIII. EXPERIMENTAL RESULTS

This section presents the experimental results obtained throughout the research. Rather than reporting only the final model, the results are presented in the same chronological sequence as the investigation evolved. This approach highlights how successive refinements in data acquisition, feature engineering, and model design contributed to the final predictive performance.

The experiments were designed to answer four principal research questions: (1) can conventional OSS counters alone predict uplink interference? (2) do atmospheric variables contain meaningful predictive information? (3) how much does geographical location contribute to prediction? and (4) which machine learning algorithm provides the best overall performance?

A. Baseline Model: OSS Features Only

The first experiment evaluated whether traditional network performance counters could adequately explain uplink interference. The model was trained using only the OSS-derived variables: UL PRB utilization and uplink traffic (L.Thrp.bits.UL). These variables represent the traffic load imposed on the network and are commonly used by radio optimization engineers when investigating uplink interference.

Contrary to expectations, the resulting prediction accuracy was very poor. The Random Forest model achieved an R^2 of approximately 0.03, indicating that traffic load alone explained only a negligible proportion of the observed interference variation. This finding suggests that uplink interference within the study area cannot be explained solely by network utilization; instead, substantial external influences must exist.

Table VII. Performance Using OSS Features Only

Feature Set	Best Model	R^2	Observation
OSS Only	Random Forest	0.03	Traffic alone cannot explain uplink interference.

B. Weather Features Only

The second experiment investigated whether atmospheric variables alone possess predictive capability. The feature set included temperature, relative humidity, pressure, dew point, wind speed, wind gust, temperature inversion, dN/dh, dM/dh, wind shear, humidity gradient, and ducting ratio.

Compared with the OSS-only model, prediction accuracy improved substantially. The Random Forest model achieved an R^2 of approximately 0.40, demonstrating that atmospheric conditions contain considerably more predictive information

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

than traditional OSS traffic counters. This result represented one of the most important findings of the study: although weather variables alone cannot fully explain uplink interference, they capture a significant proportion of its temporal variability.

Feature importance analysis revealed that several ducting-related variables consistently ranked among the most influential atmospheric predictors, including temperature, temperature inversion, dN/dh, dM/dh, and wind shear. This observation supports the original research hypothesis that atmospheric stability influences uplink interference.

Table VIII. Performance Using Weather Features Only

Feature Set	Best Model	R^2
Weather + Atmospheric	Random Forest	0.40

C. Combined Weather and Spatial Features

The third experiment introduced geographical coordinates into the feature set. Latitude and longitude were incorporated to capture spatial variability between LTE sites. Model performance increased significantly: the Random Forest model achieved an R^2 of approximately 0.58, confirming that geographical location contributes substantially to uplink interference prediction.

However, feature importance analysis produced an unexpected observation. Latitude and longitude immediately became the dominant predictors, with the model assigning considerably greater importance to spatial variables than to atmospheric features. This finding suggested that certain sites consistently experience higher baseline interference than others due to local geographical or environmental characteristics. Although geographical information improved prediction accuracy, it reduced the interpretability of the atmospheric variables, motivating additional experiments to better understand this spatial bias.

Table IX. Performance Using Weather and Location

Feature Set	Best Model	R^2
Weather + Location	Random Forest	0.58

D. Final Integrated Model

The final experiment combined all available information sources: OSS counters, surface weather observations, pressure-level atmospheric variables, engineered ducting indicators, and geographical coordinates. This integrated feature set produced the highest prediction accuracy observed throughout the study.

Among the evaluated algorithms, Random Forest consistently outperformed the remaining models. The final model achieved an R^2 of approximately 0.75, demonstrating that combining atmospheric information with conventional network statistics substantially improves uplink interference prediction.

Table X. Performance of Final Integrated Framework

Feature Set	Best Model	R^2
OSS + Weather + Location	Random Forest	0.75

E. Comparison of Machine Learning Algorithms

Five supervised regression algorithms were evaluated using the integrated dataset. Random Forest consistently produced the highest prediction accuracy ($R^2 \approx 0.75$) while maintaining excellent robustness across multiple feature combinations. Linear Regression performed poorly throughout the study, indicating that the relationship between atmospheric conditions and uplink interference is highly nonlinear. Decision Tree models suffered from overfitting, whereas Gradient Boosting exhibited greater sensitivity to parameter selection and dataset characteristics. The superior performance of Random Forest is consistent with previous studies involving heterogeneous environmental datasets.

Table XI. Comparison of Machine Learning Algorithms

Model	R^2	Adjusted R^2	MAE
Random Forest	**0.75**	**0.75**	**1.68**
Extra Trees	0.65	0.65	2.05
Decision Tree	0.49	0.49	2.27
Gradient Boosting	0.48	0.48	2.56
Linear Regression	0.01	0.01	3.66

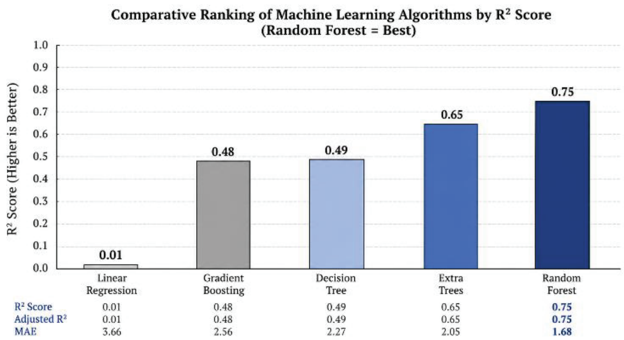

Fig. 5. Comparative ranking of the five evaluated machine learning models. Random Forest achieved the highest measured R^2 (0.75), followed by Extra Trees (0.65).

F. F ture Importance Analysis

One of the primary advantages of tree-based ensemble methods is their ability to estimate feature importance. Fig. 6 illustrates the feature importance ranking obtained from the final Random Forest model. The highest-ranking predictors were, in order: latitude, longitude, temperature, UL PRB, uplink throughput, temperature inversion, modified

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

refractivity gradient (dM/dh), and refractivity gradient (dN/dh).

The prominence of latitude and longitude indicates that geographical characteristics significantly influence the baseline interference experienced by each site. However, among the meteorological variables, temperature inversion and refractivity gradients consistently ranked above conventional weather parameters such as humidity and wind speed. This observation reinforces the importance of incorporating physically meaningful atmospheric features rather than relying solely on surface weather observations.

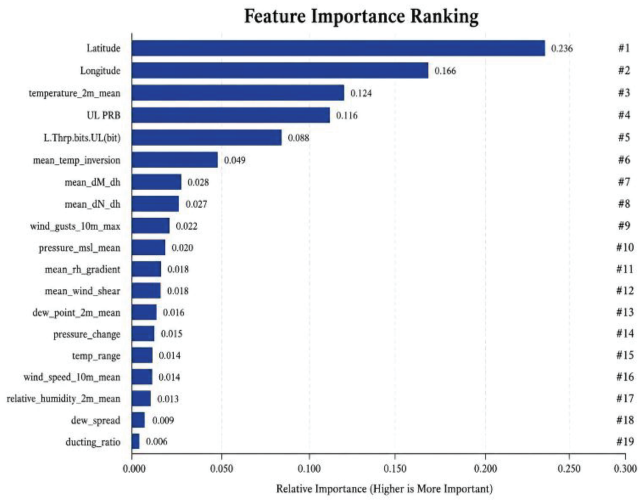

Fig. 6. Random Forest feature importance scores obtained from the final trained model. Latitude and longitude exhibit the highest importance, followed by temperature, UL PRB, uplink traffic, temperature inversion, modified refractivity gradient (dM/dh), and refractivity gradient (dN/dh).

G. Effect of Removing Spatial Features

To investigate whether the model relied excessively on geographical information, latitude and longitude were removed from the feature set. Prediction accuracy decreased noticeably. Although atmospheric variables remained predictive, the reduction in R^2 confirmed that geographical context contains valuable information.

This is largely explained by the coarse spatial resolution of the meteorological data: ERA5's 0.25° × 0.25° grid (~31 km) assigns identical weather observations to multiple nearby LTE sites that nonetheless experience different interference levels due to local network topology, terrain, and coastline proximity. Latitude and longitude therefore act as proxies for these unresolved local characteristics rather than indicating model bias; a fuller discussion of this effect and its implications is given in Section IX.C.

H. Correlation Analysis

As described in Section VI.G, Pearson correlation analysis identified and removed several redundant predictors (e.g., maximum/minimum temperature, ducting hours/ratio, correlated OSS traffic counters). Consistent with that earlier analysis, this simplification reduced model complexity with only a negligible loss in prediction accuracy, confirming that the reduced feature set retained the essential predictive information.

I. Chronological Performance Improvement

One of the distinguishing characteristics of this research is the progressive refinement of the modelling framework. Fig. 7 summarizes the improvement in prediction accuracy throughout the investigation: OSS only ($R^2 \approx 0.03$), surface weather ($R^2 \approx 0.40$), weather + location ($R^2 \approx 0.58$), and the integrated atmospheric framework ($R^2 \approx 0.75$). This progression illustrates that meaningful improvements were achieved not through algorithm changes alone, but through improved representation of the physical processes governing radio propagation.

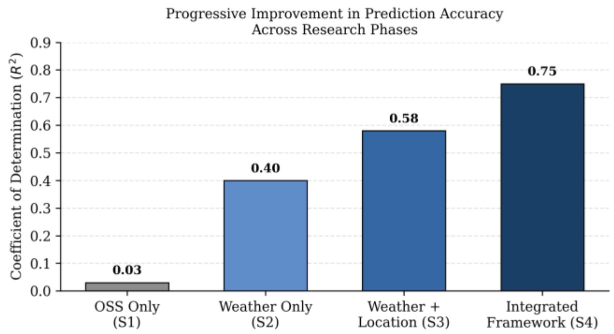

Fig. 7. Chronological improvement in prediction accuracy (R^2) across successive research phases, from OSS-only baseline to the final integrated atmospheric framework.

IX. DISCUSSION

The objective of this study was to investigate whether atmospheric conditions associated with tropospheric ducting can improve the prediction of uplink interference in a commercial LTE network. Rather than evaluating only the predictive performance of a machine learning model, the research sought to understand the physical relevance of atmospheric variables, identify the limitations of conventional OSS-based approaches, and establish a practical framework for integrating meteorological information into radio network optimization.

The experimental results demonstrate that atmospheric information provides substantial predictive value beyond traditional network performance counters. More importantly, the research highlights the importance of combining domain knowledge from radio propagation and atmospheric science with machine learning, rather than relying solely on data-driven feature selection.

A. Interpretation of Research Findings

The results confirm that uplink interference cannot be adequately explained using conventional OSS counters alone (Section VIII.A). This finding suggests that traffic loading is

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

not the dominant factor governing the observed interference behaviour within the study network, which is consistent with field experience: network engineers frequently encounter periods during which neighbouring cells experience elevated uplink interference despite stable traffic levels and unchanged radio configurations, and conventional optimization methods often struggle to explain such events because they rely almost exclusively on OSS statistics.

The inclusion of meteorological variables significantly improved predictive performance (Section VIII.B), demonstrating that environmental conditions contain valuable information regarding uplink interference variability and supporting the original hypothesis that atmospheric conditions contribute to uplink interference.

B. Importance of Atmospheric Features

One of the most significant outcomes of the research was the consistent importance of engineered atmospheric variables derived from pressure-level observations. Among all meteorological predictors, temperature, temperature inversion, modified refractivity gradient (dM/dh), refractivity gradient (dN/dh), and wind shear repeatedly appeared near the top of the Random Forest importance rankings.

Unlike conventional weather variables, these features are directly related to the physical mechanisms governing radio wave propagation. Temperature inversion is a well-established prerequisite for surface duct formation because it produces stable atmospheric layers capable of trapping electromagnetic energy. Similarly, modified refractivity gradients have long been used within propagation engineering to characterize abnormal propagation conditions.

The emergence of these variables among the most influential predictors provides evidence that the machine learning model is capturing meaningful atmospheric relationships rather than merely exploiting statistical correlations. Although the study does not prove that atmospheric ducting is the sole cause of uplink interference, the results strongly suggest that ducting-related atmospheric conditions influence interference behaviour within the investigated network.

C. Role of Geographical Location

One of the most notable findings of this study was the consistently high importance of geographical coordinates. Feature importance analysis ranked latitude and longitude among the strongest predictors, often exceeding the contribution of individual atmospheric variables. At first, this appeared counterintuitive because the objective was to predict uplink interference using weather-related features rather than site location. However, subsequent experiments showed that removing latitude and longitude resulted in a substantial reduction in prediction accuracy.

A key reason for this behavior lies in the spatial resolution of the meteorological dataset. ERA5 provides atmospheric variables on a 0.25° × 0.25° grid (approximately 31 km horizontal resolution), meaning that all LTE sites located within the same grid cell receive identical weather observations and derived atmospheric features. In practice, multiple geographically separated sites often share the same ERA5 weather data while exhibiting significantly different uplink interference levels. Consequently, the weather variables alone cannot uniquely characterize the local propagation environment of each site.

Latitude and longitude therefore provide essential spatial context by distinguishing sites that share identical atmospheric inputs. They implicitly capture location-dependent characteristics that are not resolved by the ERA5 grid, including proximity to the coastline, terrain, local clutter, antenna deployment, surrounding radio environment, and network topology. Rather than acting merely as site identifiers, geographical coordinates compensate for the limited spatial granularity of the meteorological data and enable the model to account for persistent spatial variations in baseline interference.

The final modelling framework therefore retained latitude and longitude alongside atmospheric variables, allowing the model to learn both the dynamic temporal effects associated with changing atmospheric conditions and the stable spatial characteristics associated with individual site locations.

D. Importance of Feature Engineering

Consistent with the progression described in Section V, improvements in predictive performance were driven primarily by improved feature representation rather than by increasingly complex algorithms: the earliest surface-weather-only experiments lacked the vertical atmospheric structure needed to describe tropospheric ducting, whereas introducing pressure-level observations enabled physically meaningful quantities (refractivity, modified refractivity, temperature inversion, humidity gradients, wind shear) that consistently improved both performance and interpretability. This highlights a broader principle for telecommunications machine learning: incorporating domain-specific physics into feature engineering often yields greater gains than swapping one algorithm for another.

E. Comparison with Previous Studies

Previous machine learning studies addressing cellular network optimization have predominantly relied on OSS-derived performance counters for tasks such as throughput estimation, traffic forecasting, and anomaly detection; where weather has been considered, most published work uses only surface observations rather than pressure-level, propagation-related indicators. The present study extends this literature by explicitly modelling atmospheric propagation processes through refractivity gradients and temperature inversion rather than treating weather as a generic external variable.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

F. Practical Implications for Mobile Network Operators

Uplink interference is traditionally investigated only after customer complaints or KPI degradation become visible in OSS dashboards. By incorporating atmospheric observations or short-term forecasts into predictive models, operators could instead anticipate periods of elevated interference risk — supporting proactive monitoring, early warning systems, adaptive uplink parameter tuning, and future SON functionalities — before widespread customer impact occurs. Additional validation is required before operational deployment, but the framework demonstrates the feasibility of this approach.

G. Lessons Learned from the Research Journey

Beyond the predictive framework itself, documenting the unsuccessful experiments and methodological redesigns described in Section V is intended as a practical contribution: it illustrates concretely why surface weather alone was insufficient, why spatial information mattered more than expected, and why adequate temporal coverage and physics-informed feature engineering — more so than algorithm sophistication — were the decisive factors in this investigation.

Fig. 8 summarizes how this framework could be integrated into an operational deployment pipeline, from weather forecast ingestion through to proactive SON-driven optimization.

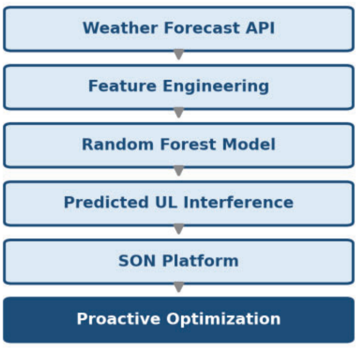

Fig. 8. Proposed deployment framework integrating weather forecasts with the SON platform for proactive uplink interference optimization.

X. LIMITATIONS

Although the proposed framework achieved encouraging predictive performance, several limitations should be acknowledged.

A. Spatial Resolution

Weather observations were assigned using the nearest ERA5 grid point. Although this approach is computationally efficient, multiple neighbouring LTE sites occasionally shared identical weather observations. Higher-resolution atmospheric datasets could potentially improve spatial representation.

B. Site-Specific Characteristics

Latitude and longitude captured substantial predictive information, suggesting that additional site-specific variables remain unmodelled. Future studies may incorporate antenna height, azimuth, electrical tilt, surrounding terrain, clutter classification, distance from coastline, and building density. Including these variables may further improve prediction accuracy while reducing reliance on geographical coordinates.

C. Missing External Interference Sources

The present study focuses primarily on atmospheric influences. Other potential causes of uplink interference, including illegal transmitters, hardware faults, industrial RF sources, and transient electromagnetic events, were not explicitly represented within the dataset. These factors may explain some of the residual prediction error observed during extreme interference events.

D. Generalization

The proposed methodology was developed using data from a commercial LTE network deployed within the Eastern Province of Saudi Arabia. Although the underlying methodology is broadly applicable, additional validation is required before generalizing the model to inland deployments, mountainous regions, different climatic zones, and other mobile network operators.

XI. FUTURE WORK

Several opportunities exist for extending the present research. First, future investigations should incorporate multiple years of atmospheric observations to improve seasonal modelling and evaluate long-term prediction stability. Second, higher-resolution numerical weather prediction datasets could provide improved representation of local atmospheric variability. Third, additional propagation-related variables such as evaporation duct height, atmospheric stability indices, and radiosonde observations could be incorporated into the feature engineering process.

Fourth, deep learning architectures including Long Short-Term Memory (LSTM) networks, Temporal Convolutional Networks (TCNs), and Transformer-based time-series models may better capture temporal dependencies within atmospheric and network data. Given that uplink interference is influenced by network topology and inter-site relationships, Graph Neural Networks (GNNs) also represent a promising direction, as they can explicitly model spatial dependencies between neighbouring LTE sites rather than treating each site independently.

Fifth, explainable artificial intelligence (XAI) techniques such as SHAP (SHapley Additive Explanations) or partial

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 07 , July - 2026

dependence analysis could provide more detailed interpretation of feature contributions, increasing confidence in operational deployment. Finally, integrating short-term weather forecasts with the proposed framework could enable predictive interference alerts several hours or days in advance, supporting proactive radio optimization and enhancing future self-organizing network capabilities.

XII. CONCLUSION

This paper presented a machine learning framework for predicting uplink interference in a commercial LTE network by integrating conventional OSS statistics with atmospheric and ducting-related meteorological features derived from pressure-level data (temperature inversion, refractivity and modified refractivity gradients, humidity gradients, wind shear, and ducting metrics). Rather than presenting only the final successful methodology, the study documented the complete research process, from an initial hypothesis linking coastal tropospheric ducting to uplink interference, through unsuccessful surface-weather-only experiments, to a physics-informed atmospheric framework spatially matched to more than 1,100 LTE sites.

Among five supervised algorithms evaluated (Linear Regression, Decision Tree, Extra Trees, Gradient Boosting, and Random Forest), Random Forest consistently provided the highest and most robust prediction accuracy. Comparative experiments confirmed a clear progression in predictive power: OSS traffic counters alone explained almost none of the variability ($R^2 \approx 0.03$), atmospheric variables alone explained roughly 40% ($R^2 \approx 0.40$), adding location raised this to $R^2 \approx 0.58$, and the complete integrated feature set achieved the highest accuracy ($R^2 \approx 0.75$). Feature importance analysis showed that latitude and longitude remained strong predictors — reflecting persistent, site-specific baseline interference not resolved by the coarse ERA5 grid — while temperature inversion, modified refractivity gradient (dM/dh), refractivity gradient (dN/dh), and temperature also ranked among the most influential atmospheric predictors, consistent with established radio propagation theory and supporting the hypothesis that tropospheric ducting influences uplink interference in the investigated network.

More broadly, this research demonstrates that improvements in predictive performance were driven primarily by physics-informed feature engineering rather than by increasingly complex algorithms, and it establishes a practical, reproducible framework for combining numerical weather data with operational network analytics. The documented research journey is intended to provide practical guidance for future work at the intersection of telecommunications engineering, atmospheric science, and machine learning, contributing toward more intelligent, proactive, and environmentally aware radio network optimization.

REFERENCES

[1] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[2] T. S. Rappaport et al., Wireless Communications and Applications Above 100 GHz. Hoboken, NJ, USA: Wiley, 2019.

[3] 3GPP TS 36.300, E-UTRA and E-UTRAN Overall Description, Release 17.

[4] 3GPP TS 38.300, NR and NG-RAN Overall Description, Release 18.

[5] S. Haykin, Neural Networks and Machine Learning, 3rd ed. Pearson, 2009.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3–42, 2006.

[8] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[9] ECMWF, "ERA5 reanalysis dataset documentation."

[10] H. Hersbach et al., "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society, 2020.

[11] ITU-R P.453, The Radio Refractive Index: Its Formula and Refractivity Data.

[12] ITU-R P.452, Prediction Procedure for the Evaluation of Microwave Interference.

[13] ITU-R P.530, Propagation Data and Prediction Methods Required for the Design of Terrestrial Line-of-Sight Systems.

[14] Open-Meteo Documentation, Historical Weather API.

[15] Open-Meteo Documentation, Historical Forecast API.

[16] H. L. Van Trees, Detection, Estimation and Modulation Theory.

[17] A. Goldsmith, Wireless Communications. Cambridge, U.K.: Cambridge Univ. Press.

[18] S. Sesia, I. Toufik, and M. Baker, LTE — The UMTS Long Term Evolution.

[19] M. Gast, 802.11 Wireless Networks: The Definitive Guide.

[20] IEEE Std. 145-2013, IEEE Standard Definitions of Terms for Antennas.

APPENDIX A. LIST OF FIGURES