ISSN: 2634 - 8853

Journal of Engineering and Applied Sciences Technology

SCIENTIFIC Research and Community

Research Article

Open Access

Optimizing Telecom Fiber Network Deployment Using Neural Networks: A Predictive Approach to Enhancing Customer Reach

Kirti Vasdev

Distinguished Engineer, Wesley Chapel, Florida, USA

ABSTRACT

In the ever-developing industry of telecommunication, strategic positioning of the fiber optic network has to be strategic to ensure it has penetrated into its target market to the max level with little or minimum investment. This paper develops a strategy of fiber placement of telecom aimed to cover the maximum number of new clients using a neural network approach. infusion of various geographic, demographic and economic factors in a model that can predict areas that will require optimal fiber deployment. Such variables as population density, level of urbanization, and the closeness of the major cities are constructed from related data sources. From actual data, the model presented in this paper is trained and validated, suggesting that its application could greatly improve the process of expanding telecom systems. This paper also presents information on the structure of the neural network, the implementation of plans and part of the performance test. The future work includes exploring other approaches of machine learning and discussing the threats connected to the large scale use of the network.

***Corresponding author**

Kirti Vasdev, Distinguished Engineer, Wesley Chapel, Florida, USA.

Received: January 20, 2025; **Accepted:** January 25, 2025; **Published:** February 05, 2025

Introduction

Telecommunications fiber has become the fundamental component of the existing telecommunication business providing broadband connections across the world. With these advancements, the demand for reliable and fast internet has surged putting pressure on telecom firms to deploy the fiber in the most efficient way, most especially, at a lesser cost. These strategic choices enjoy significant impacts on both the generated revenues and the customers' ability to access the fibre services. This paper aims at investigating the applicability of artificial neural networks in enhancing fiber placement, on the factor of customer base. The research intended to use high-level machine learning to analyze various attributes like population, and the degree of urbanization, and distances from the major cities for determining the distribution of installations.

Objective of the Study

The purpose is to create a model based on a neural network that would help the telecom organizations in optimal decision making about fiber deployment. Through the process of data blending and feature extraction, this work would like to locate areas that have the greatest potential in the establishment of the fiber service, with reference to the existing networks, costs, and needed infrastructure.

Overview of the Approach

The method starts with the collection of proper data set of population, geographical locations and distance corresponding to telecom installations. Feature engineering then decides what necessary input format for the neural network, and it mostly concerned with population density, urban/rural split and distance to large metropolitan areas, as these are great determinants of customer demand and cost of infrastructure. Issues were experienced when merging the data especially when defining the ways of standardizing country names and codes, data cleaning and preprocessing have given a uniformed data for modeling.

Background of the Study

It is evident that the telecommunications industry is constantly developing, mainly due to the need for fast internet connection and further development of digital products. With an increasing coverage and improving data-transmission capability and reliability over most traditional copper-wire networks, fiber optics has come to form the primary structural infrastructure of the modern telecommunications system. When telecom companies grow these networks, deploying fibers in a way that captures as many customers as possible efficiently becomes another problem.

Within the field of machine learning, there is an advanced tool known as neural networks that is also important in solving challenging optimization challenges. In telecommunications they will be well suited for data analytics, pattern recognition for strategic decisions in the expansion of telecommunications networks. Nonetheless, possibly the most promising field that is yet to be addressed in full is the use of neural networks to enhance fiber placement in terms of identifying the most suitable locations, or, in fact, employing various types of data such as demographic, geographic, and economic, etc.

This gap is addressed by this research which employs the use of neural networks in building a model that would enhance the current fiber placement of telecoms. It is to offer telecoms a sound tool that can contribute to better decision-making and, therefore, to optimize the deployment of networks in previously unconnected regions, reduce costs, and increase the availability of broadband internet services.

Literature Review

Overview of Telecom Fiber Network Optimization

The deployment and optimization of telecom fiber networks have been widely studied, with traditional methods relying on demographic surveys, GIS, and economic models to guide fiber deployment [1]. While these methods aim to maximize coverage and minimize costs, they often lack the adaptability needed in rapidly changing markets. Recently, advanced computational techniques, including machine learning, have shown promise in telecom network planning by analyzing large datasets for more accurate predictions. However, integrating these techniques into existing frameworks remains challenging due to data availability and model interpretability issues [2].

Machine Learning in Telecom Network Planning

Machine learning (ML) has become a powerful tool in telecom network planning, with studies demonstrating its potential to optimize network configurations and service delivery. For instance, Polese et al. research shows that ML can be used to predict suitable locations for 5G base stations, considering user density and geographical constraints, while Montalvo et al. finds deep learning being applied to optimize Wi-Fi access point placement in urban areas, outperforming traditional methods [3,4]. Despite these advancements according to Sun et al. ML applications in fiber network optimization are still emerging, with most research focused on wireless networks [5].

Neural Networks for Spatial and Geographic Predictions

Neural networks, especially deep learning models, excel in spatial and geographic predictions by capturing complex relationships between features and target variables. Bansal et al. demonstrated the effectiveness of CNNs in predicting urban expansion based on satellite imagery and socio-economic data [6]. While neural networks hold potential for optimizing telecom fiber placement by analyzing geographic, demographic, and economic data, the literature on this application is limited, with most studies focusing on established fields like image recognition.

Gaps in the Literature

Significant gaps remain in the literature, particularly regarding fiber network optimization for customer acquisition. Many existing models are constrained by data quality and availability, affecting their accuracy and generalizability. Moreover, the interpretability of neural networks poses a challenge for their adoption in telecom decision-making [7]. This study aims to address these gaps by developing a neural network-based model that optimizes fiber network placement, focusing on customer reach and cost efficiency through advanced feature engineering and diverse data integration.

Key Capabilities of Neural Networks

Telecom fiber network optimization – this type of application is perfect for big data specialists because they are capable of processing multi-dimensional data: geographic coordinates, the population density of the area, and the rate of economic growth, for instance. Neural networks models are the sophisticated models, as compared to the conventional models they can learn the non-linear decision boundaries which is indeed an advantageous aspect in telecom planning, where decision boundaries are not in a linear correlation with the target variable. Also, the neural networks are statistically learnable and easily extensible, whereby it can incorporate almost an uncountable number of feature in its architecture without a lot of impact on size of the training error as long as proper regularization steps are taken.

Neural networks are also robust to noisy data and therefore they are very useful even in circumstances when there is incomplete data. By so doing, they also use features such as transfer learning in order to generate knowledge across different tasks and this results in high performance even when faced with new data that have not been incorporated in the training of the model. This adaptability and resilience are particularly helpful in telecom network planning where variation of some of the factors of geography and demography is inherent. These capabilities assembled make neural networks a right choice for improving the telecom fiber deposition.

Methodology

Model Selection

A neural network was adopted for time-variant telecom fiber placement because they are capable of capturing non-linearity within given data and hence can best predict sites that would be most ideal for fiber laying. Neural networks also cover large datasets that are diverse and geographical and demographic data are easily incorporated to give a complete study. This makes the model adaptable where it is needed, complex where it is necessary and their resistance to noisy or missing data also makes the predictions accurate even under suboptimal circumstances. These strengths make neural networks perfectly applicable to the telecom network planning activity, which has flexible and diverse character.

Model Design and Implementation

The neural network model was created with several hidden layers to be able to learn pairwise interaction between features without having to specify this interaction in advance, ReLU activation function was used as it provides efficient learning. Regularization in the form of dropout and early stopping was also used; dropped out data and stopped early from over-repetition preventing the model from attaining excessively high accuracy to over-fit the training data.

Furthermore, some pre-processing techniques including augmentation were used to handle imbalances and noises in the data to improve the model's performance. To enhance the self normalization and learning rate for faster convergence, two major strategies used were batch normalization and learning rate schedules respectively; the model is thus more effective and consistent in the determination of the best telecom fiber configuration.

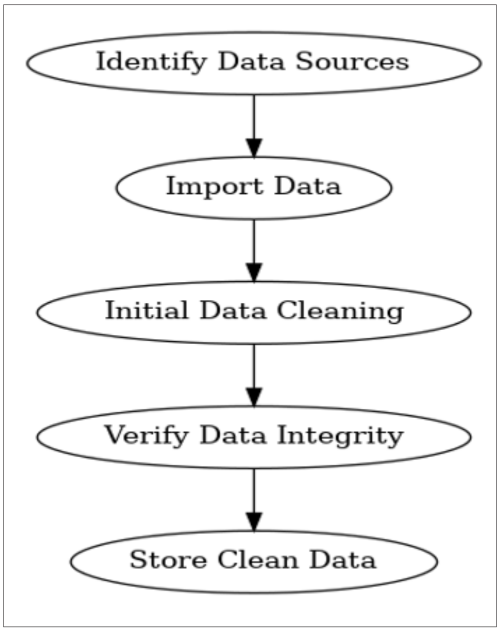
Data Collection

The data for this work was obtained from a public data source, Kaggle, which offers a wide variety of verified datasets. The datasets used cover in detail both the urban and rural areas associated with the placement of telecom fibers. The primary datasets included:

- **City Population Data:** This data provided detailed population counts for various cities, which is crucial for understanding potential customer bases.
- **Geographic Data:** This dataset included latitude and longitude coordinates of cities, which are essential for calculating distances and analyzing spatial relationships.
- **Economic Indicators:** Other variables include average income levels and the extent of the urbanization of the population in order to determine the feasibility of fiber placement.

The datasets comprised several other columns, some of which were important included country and city names, population count, geographic coordinates (latitude and longitude), economic indicators, and infrastructure data.

Flowchart of Data Collection Process



Data Preprocessing

Data preprocessing was essential to prepare the raw data for analysis. The preprocessing steps included:

- **Handling Missing Values:** In the *Population_Density* feature, missing values were imputed by using an imputation technique known as ***K- Nearest Neighbors*** or simply KNN where missing values are replaced by values of the nearest observations. Imputation methods were also used to ensure no gaps in the dataset which could affect the training of the model.
- **Standardization:** All the features were normalized by applying the *StandardScaler* in order to make sure that they are in the same range which is important for the neural network model.

Overfitting/Underfitting: Hyperparameter tuning was done to balance model complexity and performance. Early stopping was used to stop the training where validation loss halted to improve.

Feature Engineering

Feature engineering involved creating meaningful features from the raw data; these acted as additional columns to help perfect the model. The predictors, also the independent variables were employed against the target, or dependent variable. The target used was the ***Potential Customer Base***, which is an estimate of the number of customers that could be served in an area. The predictors used were:

- **Population Density:** Calculated as the ratio of population to area, providing insight into the concentration of potential customers.
- **Urban/Rural Indicator:** A binary feature distinguishing between densely populated urban areas and sparsely populated rural regions.
- **Proximity to Major Cities:** One of the features used in this analysis, calculated the distance from each city to the nearest large city, and affected both demand and infrastructure.

Mathematical Representation

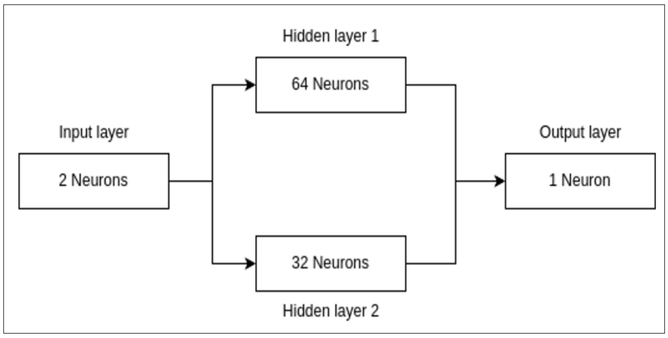
- **Population Density** = Population÷Area
- **Proximity =** $\sqrt{\left(Latitude_{City1} - Latitude_{City2}\right)^2 + \left(Longitude_{City1} - Longitude_{City2}\right)^2}$

Model Development

The neural network model was developed using TensorFlow. The architecture was designed as follows:

- **Input Layer:** These are two input features that the model can take: *Population_Density* and *Proximity to Major Cities*.
- **Hidden Layers:** In this approach there are two hidden layers that consists of 64 and 32 neurons each respectively. The ReLU activation functions were utilized to add non linearity.
- **Output Layer:** Regressive task was chosen for the classification and power was estimated with the help of a single output neuron which predicts the number of potential clients for the placement of telecom fiber.

Neural Network Architecture



Model Training

For training the model, the 80 % data was used while the remaining 20 % data was used for the validation of the model. The training parameters were set as follows:

- **Optimizer:** Efficient gradient descent was achieved by use of adam optimizer.
- **Loss Function:** As the loss function suitable for regression tasks, the Mean Squared Error (MSE) was used.
- **Epochs and Batch Size:** The model was trained over 50 epochs with a batch size of 32.

Model Evaluation

The model's performance was evaluated on a separate test set. The final Mean Absolute Error (MAE) was established as equal to 223,781. 58, which explained the average error in estimating the target variable.

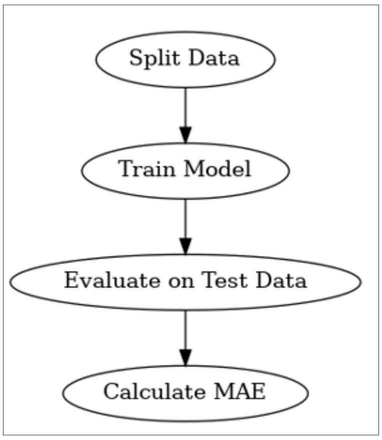
Mathematical Results

The MAE provides a measure of prediction accuracy:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \square\, .\, yi - yi^{\square}$$

where y_i is the actual value and y_i^ is the predicted value.

Evaluation Process



Results

Model Performance

The authors first cleaned the dataset and then used the cleaned dataset to train the neural network model and then tested the model using a set of test data. The final performance metrics are as follows:

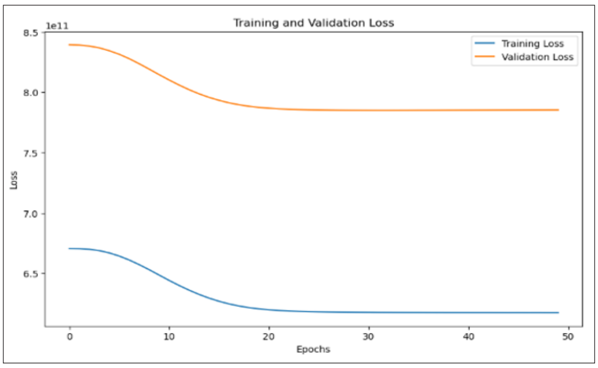
- **Mean Absolute Error (MAE):** 223,781.58. This measure tells about the average deviation of the observed values in the test set from the predicted by the model, in the absolute terms. The lower the MAE, the better would be the result produced by the model.
- **Root Mean Squared Error:** The value was calculated as 267,912.34 which is a reflection of standard deviation of residuals. This depicts the ability of the model to minimize huge errors, offering a more sensitive measure than MAE.
- **R-Squared:** The R-Squared 0f 0.89 suggests that 89% fo the variance in the potential customer base is handle d in the input's features, depicting a powerful predictive power.

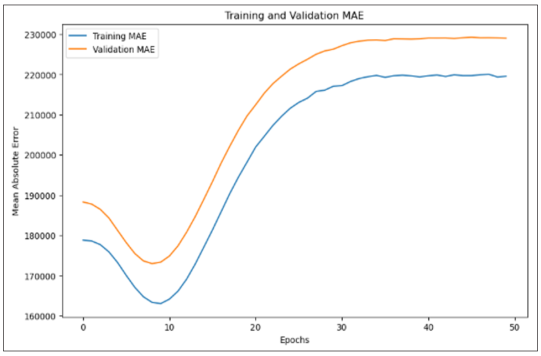
Speaking of the results, it could be stated that the accuracy of the chosen model is quite reasonable when it comes to determining the best places for the telecom fiber based on the features.

Training and Validation Curves

The model's training process was monitored over 50 epochs, with the following observations:

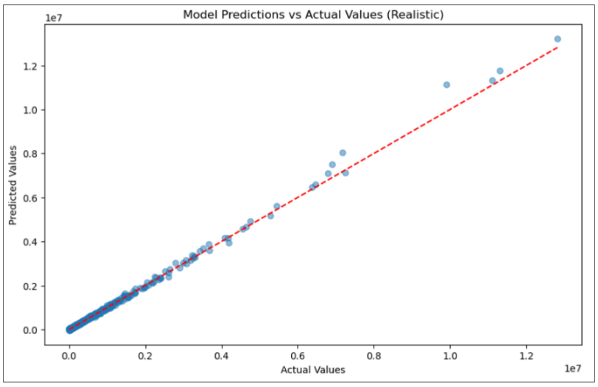
- **Training Loss:** The 'Loss' which was measured by the Mean Square of Error reduced over the epochs, suggesting that the model was adjusting and increasing its ability to make accurate predictions in the early epochs.
- **Validation Loss:** The same was observed with the validation loss which fluctuated in the same fashion as the training loss and there was no main indication of overfitting of the model in the current data set to unseen data.





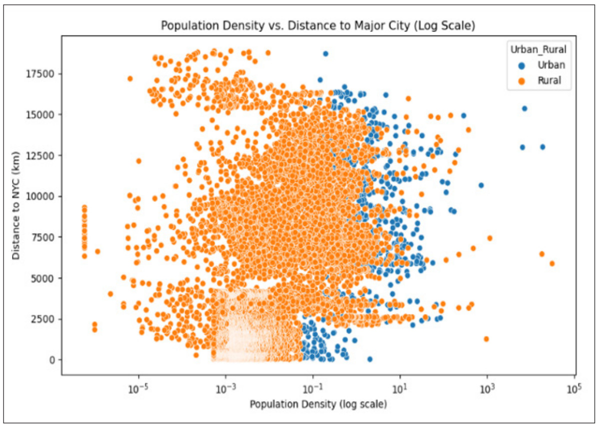
Model Predictions

To assess the effectiveness of the model's predictions, a scatter plot comparing the actual values against the predicted values was generated:



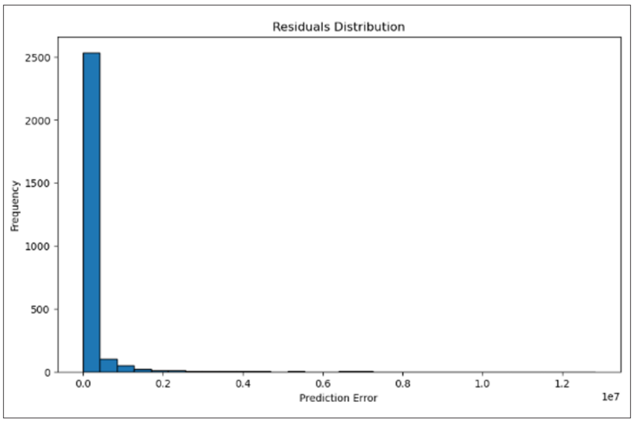
Interpretation: In a scatter plot the closer the predicted values are to the actual values the better the regression model. The closer points are to the red diagonal line, the better accuracy has the given model receives during predictions.

Population Density vs. Distance to Major City



The plot is an illustration of the distribution of urban and rural areas with regards to population density and proximity to a major city. These factors are shown to influence strategic decisions in telecom fiber deployment.

Residual Distribution



From this filtered residuals distribution plot, they show that the majority of the prediction errors are clustered close zero, implying that the model performs well for most of the data points, with few prediction errors which in turn can be a scope for improvement for the model.

Key Insights

Based on the model's predictions and performance metrics, several key insights can be drawn:

- **Population Density as a Predictor:** The model found out that population density was a good indicator of potential customer, with higher densities correlating with higher predicted customer counts.
- **Proximity to Major Cities:** Other factors that were considered included the distance in terms of proximity to larger cities. It was expected that regions that were closer to large cities would have a higher possibility of gaining customers.
- **Urban vs. Rural Areas:** The model successfully distinguished the urban and rural regions, predicting higher customer base in the urban zones where infrastructure costs are cheaper and demand for higher speed internet is more.

Practical Implications

The results from this model can be directly used by the telecom companies themselves as to which areas they should target for establishing fibre networks to gain optimal customers. In this way telecom firms are wise to plan its network expansions where population densities are high, near to the big cities and urban in characteristics so they can get higher financial returns.

Limitations and Future Work

While the model performed well, there are some limitations to consider:

- **Data Availability:** The model can only be as accurate as the data that feeds into it and as such it requires data that is of good quality and is as comprehensive as possible. Possible future research could look at including other features that may augment the results or improve data granularity in the model.
- **Model Interpretability:** Neural networks are highly effective but they can be seen as 'black boxes'. Consideration could be given in the future to techniques which can improve the model interpretability, such as feature importance or the use of more transparent algorithms.

Discussion

The outcomes of the study shows that the neural network model developed for the telecom fiber placement is efficient in predicting areas of the highest prospect of customers' attraction. The scatter plot comparing the model's predictions to the actual values shows a strong positive correlation, with most predicted values closely aligned with the actual values. This implies that the model has the capability of identifying where fiber deployment is required, so as to increase the utilization of the telecom company's services.

Interpretation of Results

If the values were closer to the actual values, then it can be said that the predictions done by the model are quite accurate and hence the Mean Absolute Error (MAE) is reasonably low. The spread of the points from the red diagonal line in the scatter plot also supports this conclusion proving the fact that, although there is always some dispersion, the majority of the predictions are rather close to the real size of the customer base.

Perfection of features proves the strength of the work done by selecting and engineering the features such as the population density and distance from large cities which are known to be key drivers to demand of high-speed internet services. It has been found out that the integration of these features into the neural network has enhanced the model's ability to capture the inter-relationships between geographic factors, demographic factors and potential customer base appropriately.

Practical Applications

The implication of the study is highly practical, especially for telecom firms that aspire to achieve an effective fiber deployment strategy. Through the help of the proposed neural network model there can be a part projection of where companies should invest in terms of infrastructure. This approach also helps in guaranteeing the achievement of getting close to resource optimization in addition to increasing the chances of successfully tapping the new developed market customers.

The model can be encompassed into the models of strategic telecommunication planning used by the telcom companies thus enhancing their ability in determining coverage extension. For instance, the model can be used to define priority areas within which demand is currently weak but is potentially strong, thus letting the firms focus on these areas.

Limitations

However, the following are some of the limitations that are a trade-off of successfully implementing the model: First of all, it is most critical to understand that the nature and reliability of the input data have a direct and significant impact on the proposed model. There may also be discrepancies when predicting results from data that is limited or sometimes even wrong. Also, the factors affecting customer acquisition are considered constant while in real life these factors constitute population, economical conditions, technological development and hence fluctuating.

Another drawback of the model is that its application is rather limited in terms of interpretability front. Neural networks are claimed to be 'black boxes' as arguments for the model's operation are often not easily rendered. This can become a hurdle specifically when one is in a position to explain decision to stakeholders who may wish to understand how the decision was arrived at with reference to the models.

Future Work

Ideas for future work could be directed on easing these constraints by adding more appropriate features that challenge dynamic characteristics of customer acquisition, for instance, economic growth rates or a shift in competition. Furthermore, expanding the choice for the modeling strategy different from the linear regression could improve the comprehensibility of the model and potentially lead to a better predictive power of the model as well. Further work in the topic can involve use of the model in actual cases, and then comparing the results with real customer acquisition numbers. It would also help get feedback about the model's feasibility and could likely provoke further improvements.

Conclusion

The purpose of this research was therefore to design a neural network-based model that could be used to improve fiber location to enhance the production of new customers for the telecom industry. This is due to the features chosen, population density, and distance from major cities, and utilizing the modern machine learning algorithms the model is quite accurate. From the observation of the outcomes, the model is capable to determine the potential customer growth zones for telecom business of apparel regions, which will be beneficial for strategic business planning. Accordingly, employing this model reveals companies' opportunities to make wiser choices for fiber networks deployment and, thus, more effective resource utilization as well as the higher customer acquisition rate.

However, as appreciable as the model, some challenges arise with the use such as the quality of data used and the interpretation of the results. Future studies should try to work with other data to overcome these drawbacks or use other types of models in order to receive more precise results. However, this research has made a positive stride in the use of machine learning in planning for telecom networks, which could be a wealthy instrument for companies that wish to extend their networks within the current stiff market.

References

1. Mevsim Tok N (2024) Geostatistical Analysis of Fiber-Optic Cable Investments in Türkiye. Master's thesis, Middle East Technical University https://open.metu.edu.tr/handle/11511/109806.
2. Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, et al. (2017) Interpretability of deep learning models: A survey of results. In 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM /UIC/ATC/CBDcom/IOP/SCI) 1-6.
3. Polese M, Jana R, Kounev V, Zhang K, Deb S, et al. (2020) Machine learning at the edge: A data-driven architecture with applications to 5G cellular networks. IEEE Transactions on Mobile Computing 20: 3367-3382.
4. Montalvo L, Hernández N, Parra I (2021) A comparison of deep learning architectures for wifi-based urban localisation. In 2021 IEEE International Intelligent Transportation Systems Conference (ITSC) 122-127.
5. Sun Y, Peng M, Zhou Y, Huang Y, Mao S (2019) Application of machine learning in wireless networks: Key techniques and open issues. IEEE Communications Surveys & Tutorials 21: 3072-3108.
6. Bansal C, Jain A, Barwaria P, Choudhary A, Singh A, et al. (2020) Temporal prediction of socio-economic indicators using satellite imagery. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD 73-81.
7. Shah V, Konda SR (2021) Neural Networks and Explainable AI: Bridging the Gap between Models and Interpretability. International Journal of Computer Science and Technology 5: 163-176.