Optimizing interconnection call routing: a machine learning approach for cost and quality efficiency

Ivy Anesu Mudari, Mainford Mutandavari, Kenneth Chiworera
School of Information Science and Technology, Harare Institute of Technology, Harare, Zimbabwe

Article Info

Article history:

Received Jun 8, 2025
Revised Sep 1, 2025
Accepted Nov 28, 2025

Keywords:

Automated
Call-routing
Interconnection
Least cost routing
Machine learning

ABSTRACT

This study presents the design and development of an automated least cost routing (LCR) model for telecommunications interconnection calls using machine learning. Leveraging a random forest regressor, the model predicts the most cost-effective call routing path based on pricing and network latency. Trained on real-world call detail records (CDRs) from TelOne Zimbabwe, the model achieved a high R^2 score of 0.851, with a mean absolute error (MAE) of $0.0482 per minute. Evaluation results demonstrate an average cost reduction of 46.75% compared to traditional routing methods, with prediction times under 0.1 seconds and latency remaining within acceptable thresholds. This work provides a practical, scalable, and efficient solution for telecom. operators seeking to reduce interconnection costs and maintain service quality through intelligent routing automation. The model architecture and performance to make it viable for integration into real-time telecom infrastructure.

Corresponding Author:

Ivy Anesu Mudari
School of Information Science and Technology, Harare Institute of Technology
Harare, Zimbabwe
Email: h230942f@hit.ac.zw

1. INTRODUCTION

Did you know that telecommunications companies can lose millions of dollars a year because interconnection calls are routed poorly? Interconnection calls, which go through two or more telecom networks, are becoming more important to the global telecommunications infrastructure. As international communication grows and digital ecosystems become more complex, the price of these calls can change a lot depending on where the call is made, government rules, and agreements between operators [1].

For telecom operators, maintaining high-quality service while controlling operational expenditures remains a persistent challenge. A critical component of this is least cost routing (LCR)—a method that selects the most cost-effective path for call termination [2]. Traditionally, routing decisions are made using static routing tables or fixed configurations. In this baseline, routing decisions are made using static routing tables, configured manually based on the lowest published tariffs available at the time of setup. These tables typically remain fixed until updated by network engineers, often on a weekly or monthly basis. These are not suitable for environments where prices and traffic conditions change [3]. This often results in telecom networks continuing to use expensive routes even when cheaper, more efficient alternatives exist.

In traditional LCR, routing decisions are made using static routing tables stored in the switch or session border controller (SBC). Each routing table entry consists of: i) prefix match: the system checks the called number (dialed number identification service) and matches it against the longest destination prefix (e.g., +44) in the table; ii) carrier list per prefix: for each prefix, one or more carriers are listed in priority order, typically sorted by lowest advertised tariff; and iii) route selection: the switch selects the first available carrier trunk in the list and routes the call. If that carrier is unavailable (congestion, no response), it falls back to the next carrier in the sequence.

Updates are done manually weekly. This means routing decisions always follow the cheapest published tariff at the time of update, without considering real-time quality signals. Latency, or the time it takes for data to be sent, is another important factor in user satisfaction, along with cost. Long delays can make calls sound worse, which can make users unhappy and cause them to leave [4]. As a result, more and more people are interested in smart routing strategies that find a balance between cost and quality of service (QoS). While the traditional method is straightforward, it suffers from several limitations: i) lack of adaptability: routes remain unchanged even when cheaper alternatives become available due to dynamic pricing or traffic fluctuations; ii) configuration overhead: manual updates introduce delays and are prone to human error; and iii) quality blindness: routing tables prioritize tariff rates without factoring in latency, jitter, or packet loss, which can lead to degraded user experience.

Previous research has looked at LCR and traffic management separately, but not many have combined real-world interconnection data with machine learning to create routing models that can predict costs. This lack of knowledge in both the academic and business worlds opens up new possibilities for routing optimization. This paper addresses that gap by proposing a machine learning-based automated LCR model, leveraging a random forest regressor. The model utilizes historical call records, operator pricing lists, and network configurations to predict the most cost-effective interconnection routes and also considers choosing routes with minimal latency.

This paper's remaining sections are organized as follows: section 2 provides a review of relevant research on machine learning applications in telecommunications routing and LCR. The model's implementation and training procedure are described in detail in section 3. The experimental results and evaluation metrics are presented in section 4. Section 5 concludes the study and discusses its implications for future research and real-world applications.

2. RELATED WORK

The associated research indicates that the use of machine learning in routing methods is a rapidly expanding innovation. It has been proven to be both effective and highly precise. Various researchers have also presented supporting theories that reinforce these findings.

Traditionally, a routing method's routing table, which included a list of potential routes and the rules for choosing the best one for a particular connection or bandwidth-allocation request, served as its fundamental tool. These rules were used by the originating node, which was in charge of making requests, to find an admissible route that satisfied the request's specifications [5]. The routing tables were configured manually, so if there was human error in configuration, it could impact the incorrect or inefficient routing. Manual updates were also costly and time-consuming, and the method lacked adaptability to dynamic network conditions.

Amin *et al.* [6] looked at how three main types of machine learning—reinforcement learning, unsupervised learning, and supervised learning—can help improve routing in software-defined networking (SDN). Their analysis led them to the conclusion that there has been a notable increase in the application of machine learning, and more especially deep reinforcement learning, for SDN routing optimization. They credited the use of machine learning for routing efficiency.

Machine learning technology was introduced in Dou *et al.* [7]. This was followed by routing algorithms that made use of different machine learning technologies. Santos *et al.* [8] demonstrated the significant potential of machine learning techniques in improving networking tasks. The ability to learn from historical datasets and apply that knowledge contributed to better outcomes. For future work, the study planned to incorporate QoS criteria into the decision-making process to ensure application-specific requirements. Additionally, the routing strategy was intended to be extended to other machine learning algorithms, with a detailed comparison among them. Gelenbe [9] also introduced the subject in network routing.

The authors concentrated on investigating reinforced learning-based and supervised learning-based routing algorithms [10], [11]. Machine learning-based routing methods were more adaptable than traditional routing algorithms in complex and dynamic network scenarios [12]. Supervised learning (including random forest regression) performed well on traffic prediction.

Using random forest regression, Zhao [5] demonstrated the effectiveness of its use for improved accuracy and stability of logistics cost prediction and provided more effective support for cost control and operation optimization of a logistics enterprise. Benefits of using the random forest regressor wereautomatic feature selection, efficient overfitting prevention, and the capacity to handle high-dimensional data [6]. Compared to Lasso regression with an R^2 of 0.39, random forest regression proved to outperform with an R^2 of 0.86. The author concluded that the random forest regressor was highly efficient for predicting the cost.

Boutaba *et al*. [13] compiled a wealth of data on machine learning techniques to assist networking. They added valuable information about the methods used, their advantages, disadvantages, and viability in actual networking situations to the conversation. Neural networks, which are used in predicting future traffic based on past traffic data, were found to be very accurate for both short-term and long-term predictions while being simple and requiring few features. The paper "optimal network route estimator using prediction algorithms", proposed a prediction-based routing model that identifies optimal network paths to support efficient data transmission from source to destination IP addresses, thereby improving overall transmission efficiency in terms of quality-of-service metrics [14].

Hailan and Alshaheen [15] reviewed and examined the concepts of network planning and optimization, traffic engineering, and QoS to find practical ways to improve network performance and reduce latency and packet loss. The study emphasized the importance of advanced traffic engineering strategies to manage the rising network complexity and the growing demand for high-speed connectivity and diverse services. It identified key directions for future research, including the development of dynamic and adaptive traffic management approaches leveraging machine learning.

A study to examine how service quality, trust, and commitment influence customer loyalty among telecom service users in India, while also assessing the moderating effects of gender, marital status, and connection type. Using survey responses, the study found that responsiveness, assurance, and empathy positively affect both trust and commitment, whereas tangibility influences trust only. Commitment and trust were shown to positively impact loyalty [16].

The literature illustrated the wide-ranging use of different approaches and techniques to impede network routing optimization. It made use of the potential for machine learning to bridge the gap caused by traditional manual approaches. Researchers found that supervised learning algorithms were the most effective in these optimization settings. The random forest regressor also proved to have high prediction accuracy with minimal error. The research did not incorporate real-world telecom data integration.

3. METHOD

This study will concentrate on creating and assessing a machine learning-based least-cost routing model that takes into account crucial factors like call costs and network conditions using real-world telecom datasets. The purpose of the study is to automate call routing whilst minimizing call routing costs with enhanced network utilization. The effectiveness of the model will be confirmed through comparison with traditional routing techniques. Telecom operators will gain a competitive edge with the suggested model because it allows them to attract and retain consumers by providing high service reliability and reducing interconnection costs. i) to draw in and keep consumers, businesses must provide high service reliability and reduced interconnection costs; ii) telecom companies can more efficiently allocate resources thanks to automation, which lowers operational overhead; and iii) by cutting expenses, increasing productivity, and guaranteeing the best possible call quality for customers, a machine learning-powered system gives telecom operators a competitive advantage.

How the machine learning model fits within a telecom routing stack can be seen in Figure 1. Using an applied quantitative research design, this work aims to develop and validate a machine learning model that can predict the best interconnection routes for voice calls. To optimize for call cost and latency, the study focuses on creating a random forest regressor model. TelOne Zimbabwe provides historical telecom connectivity data for training and evaluating the model.

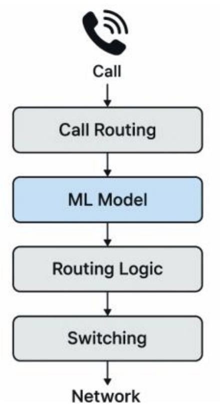


Figure 1. ML model integration within the telecom call routing stack

3.1. Planning and data preparation

The process of designing and developing a research conceptual framework involved gathering the required call detail records (CDRs), data transformation, and analyzing the data for the model. The CDRs were readily available at no cost. No human subjects were directly involved in this research. We ensured ethical clearance and privacy compliance by including historical interconnection call logs in the dataset.

3.2. Data and pre-processing

The data and pre-processing stage involve several steps to prepare the CDRs for model development, as outlined follows:

i) CDR data collection: the process of collecting CDR data samples from the database of previously captured CDR files was done through the Department of TelOne Interconnection. The data has the following columns:
   Dataset Columns
   calling_number
   called_number
   call_duration_min
   originating_carrier
   destination_carrier
   destination_country
   route_used
   cost_per_min_usd
   latency_ms
   is_peak_hour
   total_cost_usd
   day_of_week
   call_hour
ii) Anonymization: in compliance with the data protection act [17], caller numbers were converted into new numbers before use. This is due to the extreme sensitivity of the telecommunications data [18], [19]. This also helped to separate identities of customers' and their personal business.

3.3. Feature selection

The following the selection and extraction of significant features, several features were identified. These features were found to be the most pertinent in determining call routing costs. The results are shown in Table 1.

Table 1. Data fields and their relevance for prediction

Field	Relevance
originating_carrier	The originating_carrier helps identify the initial network, which influences routing decisions.
destination_carrier	Destination_carrier serves as an indicator of the target network, influencing pricing and interconnection agreements.
destination_country	Essential for figuring out costs, rules, and routing routes.
route_used	The display of the current routing path is essential for training and validation purposes.
cost_per_min_usd	The objectives of LCR optimization are closely related to cost.
latency_ms	The need to balance cost and quality may influence route selection.
is_peak_hour	Network congestion may impact both available routes and costs.
day_of_week	It may impact call patterns but has less of a direct cost impact.
call_hour	The impact on network load and pricing models makes it moderately relevant.

3.4. Materials and tools

The materials and tools used in this study include the Python for programming language. The libraries used are Scikit-learn, Pandas, NumPy, and Matplotlib, along with the random forest regressor for the model. The platform used is a local machine with sufficient computational power for training.

3.5. Model development

The main objective is to predict the carrier or route with the lowest projected cost while meeting QoS requirements. To achieve this, a supervised learning algorithm was used to develop a predictive model that can identify patterns in data, forecast future trends, and make fact-based decisions [20]. The random forest regressor was trained to forecast the most economical path. The proposed framework for general model training and logic programming was developed in Python. The model was trained to minimize mean absolute error (MAE) and mean squared error (MSE), predicting the lowest-cost route given a set of input conditions.

3.5.1. Random forest regressor

The algorithm uses the mathematical equation shown as (1).

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad (1)$$

Where, ŷ is predicted cost, N is number of trees, and $T_i(x)$ is prediction from tree i for input x.

Using various data subsets, random forests construct multiple decision trees before averaging the outcomes. The most influential features affecting call costs can be identified using random forests to calculate feature importance scores [21], [22], as illustrated in Figure 2. This benefits business insights as well as model transparency. One method that can reduce overfitting is random forests [23], [24]. Because it uses fewer parameters than other ensemble algorithms to process large datasets, it is also quick and efficient [25]. The formula was used to determine the average latency experienced when utilizing the chosen route, as well as potential routes that might provide lower call termination rates.

The model's performance was validated, and its generalizability was ensured using a 70/30 train-test split and 5-fold cross-validation. Random forest was chosen for its ability to balance strong predictive performance with interpretability, making it well-suited for applications where understanding model decisions is important. The study gained practical significance, and the model's validity was improved by using real-world telecom data from an operational setting, which reflected actual routing and cost conditions. To make sure the model included the most pertinent elements affecting call cost and quality, feature selection was guided by domain experience and concentrated on variables strongly related to routing economics and network restrictions.

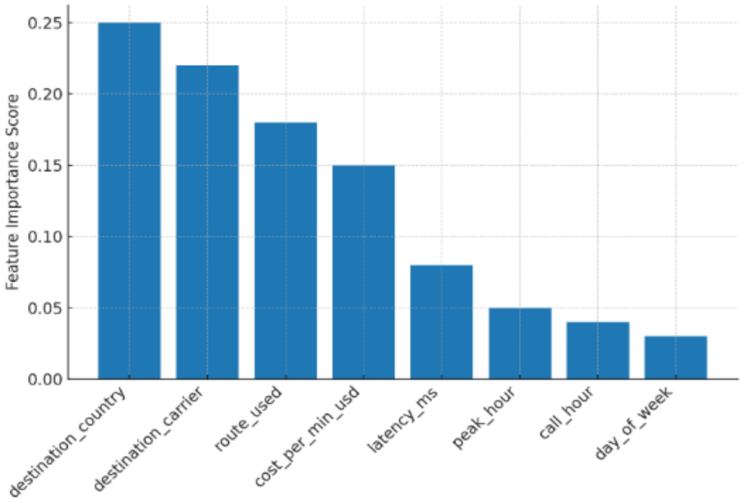


Figure 2. Random forest feature importance for least-cost routing

3.6. Evaluation metrics

To determine the distinction and importance, the model's precision and effectiveness in forecasting the best path were contrasted with conventional routing techniques. Key performance indicators such as acceptable latency, least-cost route identification, and prediction time were analyzed. The following evaluation metrics were measured:

i) Cost performance: MAE and MSE, which represent the discrepancy between expected and actual costs.
ii) Latency comparison across predicted versus traditional routes (threshold 300 ms).
iii) Prediction time: real-time feasibility was evaluated by accounting for the average time taken by the model to predict the optimal route.
iv) Operational efficiency: a comparison of the overall cost savings from the machine learning model with the traditional LCR techniques, which relied on static tables.

The study's conclusions may not be as broadly applicable to other networks or geographical areas because it is based on interconnection data from a single telecom operator, TelOne. Furthermore, because the model is trained using historical pricing data, it may become less successful in environments with pricing structures that change often, requiring retraining on a regular basis. Additionally, even though latency data was captured for every conversation, it only represents the average circumstances at the time of the call and might not accurately reflect network oscillations or brief delays that occur in real time.

3.7. Overall model architecture

Optimal route prediction workflow is illustrated in Figure 3. This figure provides an example of the process that will be used to make the prediction. The workflow also highlights how each stage contributes to generating the final routing output.

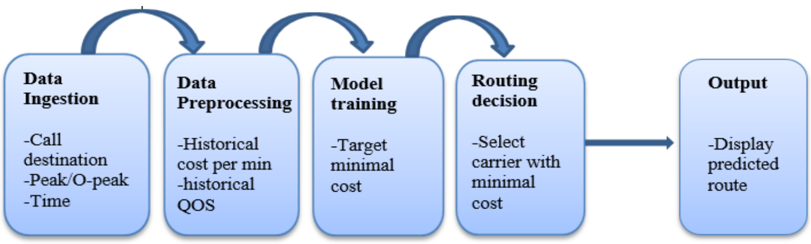


Figure 3. Overall workflow of the optimal route prediction model

4. RESULTS AND DISCUSSION

4.1. Model training results

The trained model shows strong performance, achieving an R^2 of 0.851. It also produces a low MAE of $0.0482 per minute, indicating small average prediction errors. Additionally, the MSE of 0.0032 confirms that the model provides reliable and accurate estimates.

4.2. Overall model results

4.2.1. Route selection (input)

The route selection process requires several input parameters that influence the model's decision. These inputs play a crucial role in determining the optimal route for each call. As shown in Figure 4, the interface allows users to specify the destination country, peak/off-peak, day of week, and hour of day.

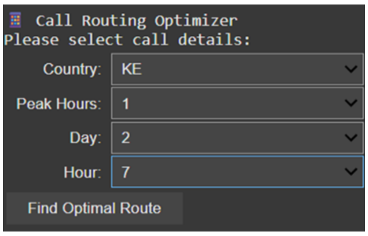


Figure 4. User interface for selecting call routing input parameters

4.2.2. Routing recommendation results (output)

The model produces several output parameters that describe the recommended route and its performance characteristics These outputs provide insight into both the efficiency and reliability of the selected route. As illustrated in Figure 5, the system displays the recommended route along with key metrics such as route carriers, latency, predicted cost, success rate, and route ID.

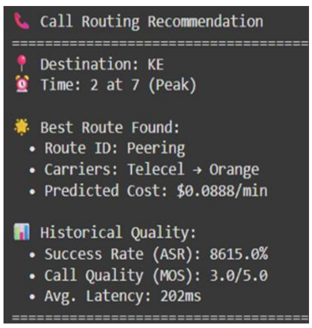


Figure 5. Model output displaying the recommended route and its performance metrics

4.3. Prediction time

The prediction time evaluation measures how fast the model can generate a routing decision. As shown in Figure 6, the Python code snippet calculates the average prediction time across multiple runs. The results indicate that the model can predict the optimal route in an average of 0.080993 seconds.

```
import time
import numpy as np

|

n_runs = 100  # Number of runs to average
total_time = 0

for _ in range(n_runs):
    start_time = time.time()
    y_pred = model.predict(X_test)
    end_time = time.time()
    total_time += (end_time - start_time)

average_time = total_time / n_runs
print(f"Average prediction time: {average_time:.6f} seconds")
```

Figure 6. Python script used to measure the model's average prediction time

4.4. Cost comparison

The cost comparison highlights the difference between traditionally routed costs and the model-generated routing costs. As shown in Figure 7, the model consistently produces lower costs across all destinations compared to traditional routing. Furthermore, Table 2 shows that the implementation of the model results in an average cost reduction of approximately 46.75%, demonstrating its effectiveness in improving routing efficiency.

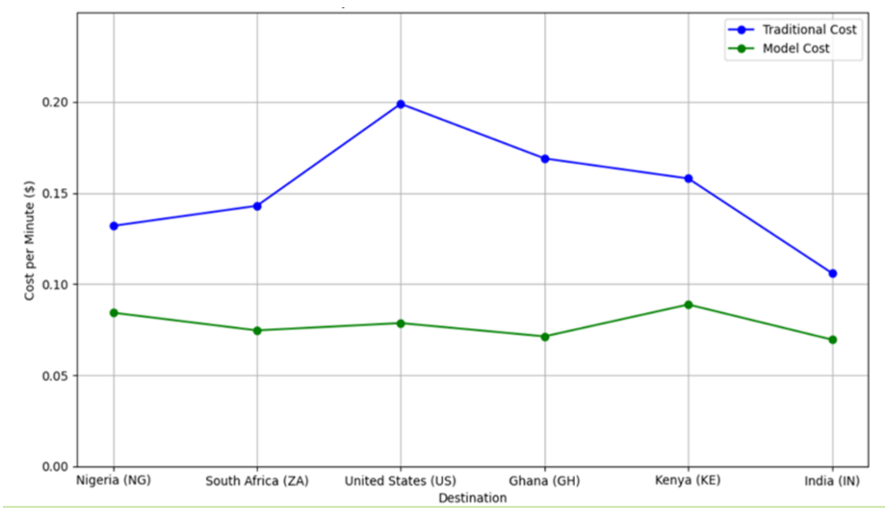


Figure 7. Comparison of traditional cost vs model cost

Table 2. Average percentage of cost difference

Destination	Route used	Traditional-cost	Model-cost	Cost difference/min	% Difference
NG (Nigeria)	Tigo-Telecel	0.132	0.0843	0.0477	36.13636
ZA (South Africa)	Telecel-Orange	0.143	0.0746	0.0684	47.83217
US (United States)	Telecel-Zamtel	0.199	0.0786	0.1204	60.50251
GH (Ghana)	Zamtel-Zamtel	0.169	0.0713	0.0977	57.81065
KE (Kenya)	Telecel-Orange	0.158	0.0888	0.0692	43.79747
IN (India)	Telecel-Voda	0.106	0.0695	0.0365	34.4339
Average % cost difference					46.75218

4.5. Latency analysis

The delay associated with each route selected for each country was evaluated to validate the efficiency of the selected optimal route, as shown in Figure 8. This analysis helps to assess how quickly the model can process routing decisions compared to traditional methods. From the figure, it was highlighted that the maximum latency was under 300 milliseconds.

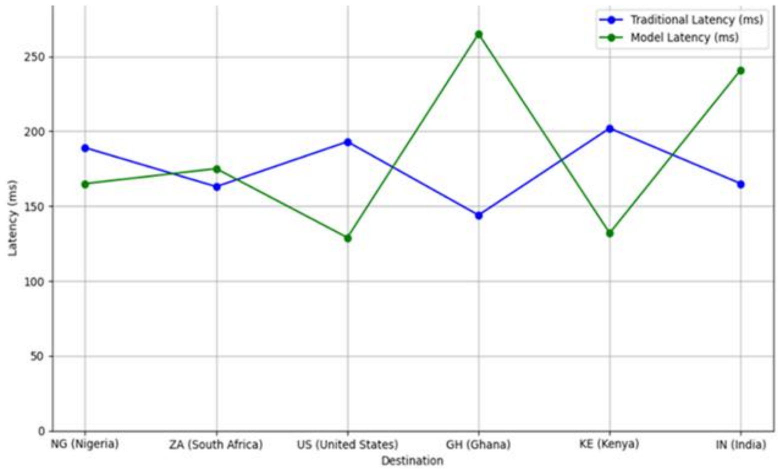


Figure 8. Comparison of traditional latency vs model latency

5.6. Discussion

The model has explanatory power, accounting for 85.1% of the variance in call cost per minute, with an average absolute deviation of 4.82 cents per minute between the model's predicted cost and the actual cost. The MSE between the model's predictions and the actual expenses is 0.0032, signifying commendable model performance. The research illustrates the efficacy of a machine learning-based approach for optimizing least-cost routing in interconnection calls. The methodology enhances routing decisions to minimize costs and ensure superior service by leveraging previous data. The machine learning methodology decreased costs by an average of 46.75% relative to conventional routing methods. Conventional routed costs (blue) exceed model-routing costs (green), with a statistically significant difference, as illustrated in Figure 7. Given that the model's latency was satisfactory, the QoS remained unimpeded, resulting in an overall equilibrium between cost and quality. The model processed routing decisions in 0.08 seconds, proving viable for real-time deployment.

The model interfaces with historical systems through Python APIs to facilitate functionality in the switch systems that govern call routing. Future real-time adaptability may investigate reinforcement learning for dynamic pricing. The study's limitations encompass reliance on the precision and comprehensiveness of historical data and the unique characteristics of pricing agreements, which may differ significantly among providers and locations.

5. CONCLUSION

This research successfully developed and evaluated a machine learning-based LCR model using real interconnection data from a major telecom operator. The random forest regressor model demonstrated strong predictive accuracy, enabling up to 60% cost savings while maintaining acceptable call quality with latency under 300 ms. With an average prediction time of just 0.08 seconds, the model proves feasible for real-time deployment. These results highlight the practical value of applying machine learning to optimize telecom routing decisions—reducing operational costs, enhancing efficiency, and supporting high service reliability. The use of historical telecom data and measurable performance metrics ensures that the solution is both robust and grounded in real-world conditions. Looking ahead, this model can be extended to multi-operator environments, integrated with 5G/VoIP systems, and enhanced using reinforcement learning for dynamic adaptation to fluctuating pricing and network conditions. Additionally, incorporating blockchain-based settlements may further streamline inter-operator cost reconciliation. Overall, this study underscores the growing relevance of AI-driven automation in telecommunications and lays the foundation for smarter, cost-efficient, and adaptive routing systems.

ACKNOWLEDGMENTS

We want to sincerely thank TelOne Pvt. Ltd. for helping with the model's requirements and for supplying the historical CDRs that enabled this study, as well as for helping with anonymizing confidential

data on the dataset guided by the law regarding customer data protection. Their knowledge, perceptions, and helpful criticism have served as the foundation for this study.

FUNDING INFORMATION

The authors state that no funding is involved in doing this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ivy Anesu Mudari	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Mainford Mutandavari		✓		✓	✓	✓		✓		✓	✓	✓	✓	
Kenneth Chiworera		✓		✓	✓	✓		✓		✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors states that no known competing financial interests or personal relationships could have appeared to have an impact on the work reported in this paper. No conflicts of interest are disclosed by the writers.

ETHICAL APPROVAL

This study complies with ethical standards for telecommunications research. All personal identifiers were removed, and all data used was transformed [24]. No private data was revealed [25]. The Harare Institute of Technology gave its institutional approval for this study.

DATA AVAILABILITY

Due to its proprietary nature and the inclusion of commercially sensitive data, the dataset's accessibility is limited. However, upon request, the author [IAM] may provide the anonymized dataset used for model evaluation.

REFERENCES

[1] H. Intven, J. Oliver, and E. Sepulveda, *Telecommunications regulation handbook*. Washington, D.C., United States: The World Bank, 2000.
[2] S. Itive, “Operational excellence strategies in the telecommunication industry,” *International Journal of Research Publication and Reviews*, vol. 5, no. 12, pp. 1022–1027, 2024.
[3] D. Mitra, S. Sarkar, D. Hati, and D. Mitra, “A comparative study of routing protocols,” *International Research Journal of Advanced Engineering and Science*, vol. 2, no. 1, pp. 46–50, 2016.
[4] I. Arapakis, S. Park, and M. Pielot, “Impact of response latency on user behaviour in mobile web search,” *arXiv:2101.09086*, 2021.
[5] S. Zhao, “Research on logistics cost prediction based on random forest regression modeling,” *Transactions on Economics, Business and Management Research*, vol. 11, pp. 367–376, Nov. 2024, doi: 10.62051/mv8p5485.
[6] R. Amin, E. Rojas, A. Aqdus, S. Ramzan, D. C.- Perez, and J. M. Arco, “A survey on machine learning techniques for routing optimization in SDN,” *IEEE Access*, vol. 9, pp. 104582–104611, 2021, doi: 10.1109/ACCESS.2021.3099092.
[7] X. Dou, Z. Wang, S. Li, and X. Zhang, “A survey on routing algorithms based on machine learning,” *Insights in Computer, Signals and Systems*, vol. 1, no. 1, pp. 79–87, 2024, doi: 10.70088/aztvzm93.
[8] C. L. D. Santos, A. M. Mezher, J. P. A. Leon, J. C. Barrera, E. C. Guerra, and J. Meng, “ML-RPL: machine learning-based routing protocol for wireless smart grid networks,” *IEEE Access*, vol. 11, pp. 57401–57414, 2023, doi: 10.1109/ACCESS.2023.3283208.
[9] E. Gelenbe, “Machine learning for network routing,” *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–1, 2020, doi: 10.1109/MECO49872.2020.9134073.
[10] C. Zhao, M. Ye, X. Xue, J. Lv, Q. Jiang, and Y. Wang, “DRL-M4MR: an intelligent multicast routing approach based on DQN deep reinforcement learning in SDN,” *Physical Communication*, vol. 55, 2022, doi: 10.1016/j.phycom.2022.101919.
[11] S. Yao, X. Lin, J. Wang, Q. Zhang, and Z. Wang, “Rethinking supervised learning-based neural combinatorial optimization for routing problem,” *ACM Transactions on Evolutionary Learning and Optimization*, 2024, doi: 10.1145/3694690.

[12] D. Turlykozhayeva *et al.*, "Evaluating machine learning-based routing algorithms on various wireless network topologies," in *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments*, Dec. 2024. doi: 10.1117/12.3058676.

[13] R. Boutaba *et al.*, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, 2018, doi: 10.1186/s13174-018-0087-2.

[14] B. Sahana, A. Agarwal, A. Ameta, A. Bhati, B. Sadhana, and A. Deshapande, "Optimal network route estimator using prediction algorithms," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 1–5, Jul. 2020, doi: 10.1109/CONECCT50063.2020.9198468.

[15] A. M. Hailan and H. Alshaheen, "Survey of traffic engineering solutions for a telecommunication network optimization," *Tikrit Journal of Engineering Sciences*, vol. 31, no. 2, pp. 128–137, 2024, doi: 10.25130/tjes.31.2.12.

[16] K. Yum and B. Yoo, "The impact of service quality on customer loyalty through customer satisfaction in mobile social media," *Sustainability*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151411214.

[17] Zimbabwean Government Gazette, *Cyber and data protection (licensing of data controllers and appointment of data protection officers) regulations*, Statutory Instrument 155 of 2024, vol. CII, no. 72, 2024.

[18] Zimbabwe Teachers Association, *Education act [chapter 25:04]*. 2016.

[19] Zimbabwe Teachers Association, *Data protection act [chapter 11:12]*. 2021.

[20] A. Bello, I. Ajibade, I. Wonuola, and D. Ekweli, "Developing a predictive model for student academic performance using machine learning techniques," *International Journal of Science and Research Archive*, vol. 16, no. 3, pp. 219–234, 2025, doi: 10.30574/ijsra.2025.16.3.2539.

[21] X. Yuan, S. Liu, W. Feng, and G. Dauphin, "Feature importance ranking of random forest-based end-to-end learning algorithm," *Remote Sensing*, vol. 15, no. 21, 2023, doi: 10.3390/rs15215203.

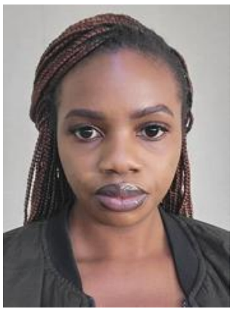
[22] G. Ma, J. Chen, J. Li, H. Shi, and Y. Chen, "Random forest regression may become the optimal regression model for osteoarthritis of the knee in elderly, in the context of embodied cognition and psychosomatic medicine," *Journal of Multidisciplinary Healthcare*, vol. 18, pp. 4219–4232, 2025, doi: 10.2147/JMDH.S519195.

[23] F. Pedregosa *et al.*, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[24] L. Barreñada, P. Dhiman, D. Timmerman, A.-L. Boulesteix, and B. Van Calster, "Understanding overfitting in random forest for probability estimation: a visualization and simulation study," *Diagnostic and Prognostic Research*, vol. 8, no. 1, 2024, doi: 10.1186/s41512-024-00177-1.

[25] H. A. Zeini, D. Al-Jeznawi, H. Imran, L. F. A. Bernardo, Z. Al-Khafaji, and K. A. Ostrowski, "Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil," *Sustainability*, vol. 15, no. 2, Jan. 2023, doi: 10.3390/su15021408.

BIOGRAPHIES OF AUTHORS



Ivy Anesu Mudari is a Bachelor of Technology holder in Software Engineering from the School of Information Science and Technology, Harare Institute of Technology. She is currently pursuing a master of technology degree in cloud computing. She also works with a telecommunications company, which has driven her to consider interconnection as an area of interest. She can be contacted at email: h230942f@hit.ac.zw.

Mainford Mutandavari is a Ph.D. scholar at SRMIST University, India, and a lecturer and postgraduate studies coordinator at the Harare Institute of Technology (HIT), Zimbabwe. With advanced degrees in computer science and strategy and innovation, his research spans data analytics, cybersecurity, IoT, AI, and cloud computing. He is a member of HIT's Cybersecurity and AI research groups and actively contributes to national ICT standards through the Standards Association of Zimbabwe. He has published widely on topics such as data loss preventions, digital learning infrastructure, and e-health security. His work bridges academic research with industry applications by focusing on practical digital solutions for education, telecommunications, and healthcare in Zimbabwe. He is also involved in curriculum development, postgraduate supervision, and building academic-industry partnerships. He can be contacted at email: mmutandavari@hit.ac.zw.

Kenneth Chiworera is a lecturer at the Harare Institute of Technology (HIT), where he coordinates the HIT200 projects and supervises undergraduate student's bachelor of technology (B.Tech.) and a master of technology (M.Tech.) in software engineering. His research interests include cloud computing, cybersecurity, artificial intelligence (AI), and blockchain technology. He is dedicated to mentoring students, fostering innovation, and enhancing practical skills through project-based learning. He actively participates in academic conferences and workshops, contributing to advancements in technology. He can be contacted at email: kchiworera@hit.ac.zw.