INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Tracing the exact location of failures in underground optical networks using LSTM deep learning model

Owusu Nyarko-Boateng[1]*, **Adebayo F Adekoya**[1], **Benjamin A Weyori**[1]

1 Department of Computer Science and Informatics, the University of Energy and Natural Resources, Sunyani, Ghana

OPEN ACCESS

Received: 04.11.2020

Accepted: 29.12.2020

Published: 02.02.2021

Citation: Nyarko-Boateng O, Adekoya AF, Weyori BA (2021) Tracing the exact location of failures in underground optical networks using LSTM deep learning model. Indian Journal of Science and Technology 14(4): 297-309. https://doi.org/10.17485/IJST/v14i4.2008

* **Corresponding author**.

owusu.nyarko-boateng@uenr.edu.gh

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Failures mostly occur in the underground optical networks, and these failures come with some consequences such as interrupted services, damaged brand reputation, revenue loss, and subscribers churn. Several fault tracing techniques in optical networks, notably OTDR, OFDR, T-OTDR, among many others have been developed, but their main challenge has been the impreciseness of failure location measurements. These challenges resulted in delays in the fault tracing processes. This paper provides a solution to the identified challenge by designing a predictive model to determine the damaged underground fiber cable's exact location. We collected datasets comprising the OTDR measurements, the number of spliced joints, the number of enclosure boxes, and the wells (chambers) along the failed underground optical transmission path. We adopted an LSTM deep learning predictive model to find the exact distance of failure in the underground optical networks. Using deep learning predictive models to tracing faults has been deployed in many network and transmission systems such as routing, fault management, link optimization, modulation, etc. However, tracing failures in underground optical networks has not yet received much attention, according to literature. Hence, the motivation to design an intelligent predictive model based on the LSTM deep learning technique to predict the exact location of failure in underground optical networks. Our study performed very well with high accuracy of 0.00016673. This result signifies that the predictive model could promptly predict the exact spot of failures in underground optical networks. This is the first time an LSTM model has been used to accurately predict the precise location of failures in underground optical networks. In addition, a computational algorithm was developed as a supporting feature to enhance the performance of the intelligent model.

Keywords: Fiber cable cuts; fault tracing; underground optical networks; LSTM; predictive model

1 Introduction

Underground fiber cable-cut is a major cause of failures in optical network infrastructure in low-income countries such as Ghana, Nigeria, Kenya, India, among many other countries. Ghana has five fiber submarine cable landing stations. These cables include Glo_1 cable, Main-one cable, West Africa Cable Service (WACS), African Coast to Europe (ACE), and South Atlantics Telecommunication-3/West Africa Cable Service (SAT-3/WACS)[1]. These cables originate from various parts of Africa with a landing station in several other nations across the coast of the continent[2]. Figure 1 shows the basic architecture of the ring topology of an underground fiber network in Ghana. The installed fiber cable from the landing station connects to the main switching centres (SC) of the mobile network operator (MNO) by underground transmission system[3].

A distributed fiber cable is then used to link the SC and other hubs (backbone) before it gets the last mile. These links deploy an underground transmission system in which the main mode long-haul transmission in Ghana[4,5]. The fiber cables buried underground sometimes develop fault according to the findings of[5]. Typical among these faults is when there is fiber cable cut within the earth. These cables cut mostly do not show up on the surface of the earth for easy fault tracing and identification. This phenomenon leads to difficulty in faults tracing in the underground transmission infrastructure[3].

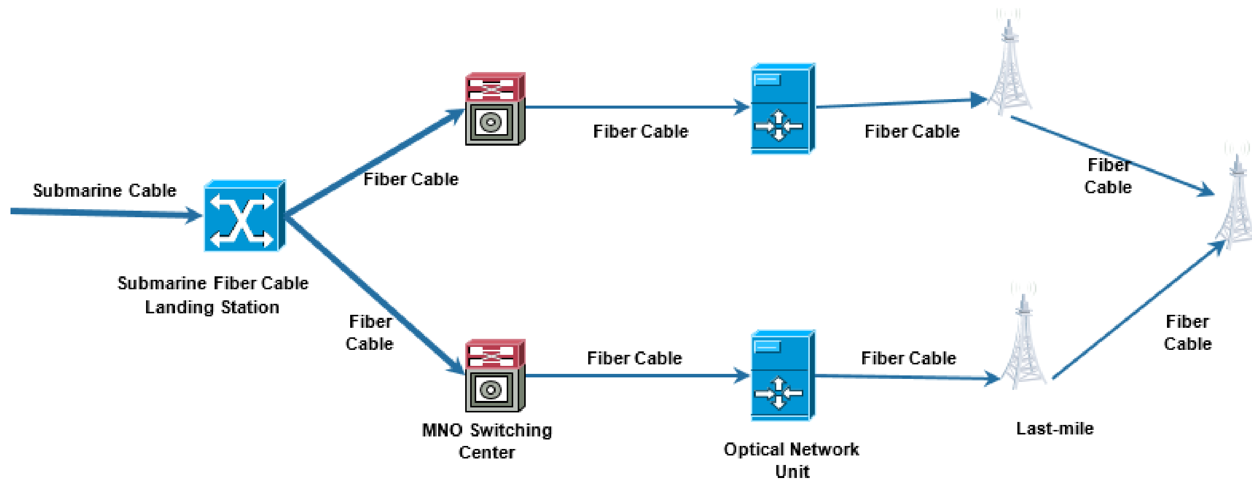

Fig 1. Physical topology of an underground optical network

When the cut occurs, the OTDR device is deployed to measure the distance of the underground cable. The use of OTDR in tracing fault distance of FOC takes much time because the device only reports the estimated length of underground fiber cable from the optical transmitter to the point of fault, so locating the fault on earth's surface is quite a difficult task. There is an excessive delay in pinpointing the exact place of fault on the earth's surface since the fault is in the fiber cable buried underground. There is an extreme waste of resources such as the cost of labour for extra digging in search of the underground fiber cable cut point[6,7].

In this research, we identify machine learning models from the literature with the potential for fault tracing in an underground transmission system, which is the long short-term memory (LSTM) networks, a recurrent type of neural network. LSTM is an architecture that includes a feedback loop mechanism delayed in time with memory cells, which enable the model to recollect past information. LSTM belongs to the deep learning, a subfield of machine learning. Given this, the main objective of this study is to conduct fault tracing simulations with the LSTM model, to assess the potential of LSTM networks in predicting the actual fault distance in the underground optical network[8,9].

Given this objective, the main contribution is the extension of existing machine learning models to the fault tracing process of the underground optical network that has not comprehensively and extensively been covered using the LSTM model[10–12]. To the best of our knowledge, this is the first study of using LSTMs for an underground optical network fault tracing.

The contributions of our paper are as follows. First, we developed an intelligent predictive model based on the LSTM model to predict the exact location where the failure in the fiber cable occurred. The intelligent model was able to accurately predict the exact location. Secondly, we developed a mathematical algorithm as an intelligent supporting model to enhance the performance of the LSTM model. This mathematical algorithm supported the model to have the best accuracy in prediction. The additional information on the stock of cables (coiled cables) in the chambers and the number of splice enclosure boxes along the fiber cable transmission path, gave us better results.

The rest of this paper has been organized as follows: In Section 2, the review of related literature was presented. In Section 3, we explain the underground optical networks and the theory of LSTM deep learning model, together with the memory cells gates in the LSTM model. Section 4 describes the methodology, the design of the study and the datasets used. The results were presented and discussed in Section 5, and finally, in Section 6, the conclusion was outlined together with direction for future work.

2 Related works

The literature review starts by covering studies focusing specifically on the use of the existing techniques of tracing failures in underground optical networks [(13)]. However, given the limited number of publications covering machine learning applications in this domain, the review was extended to other areas of machine learning application for solving problems through classification, pattern recognition, planning, and many others.

The well-known technique used to determine the distance of failure in an underground fiber network is OTDR [(14)]. The OTDR is a device that provides some properties which are essential to determine the integrity of the fiber cable in the optical network infrastructure. OTDR measures the reliability of a fiber cable, attenuation, splice loss, length of cable, micro/macro bends, and then detect a failure point in the fiber cable [(15–17)]. However, the OTDR technique faces the challenges of precision and accuracy of given the physical distance of failure in an underground optical transmission system. The OTDR uses two main working principles to compute the distance of fault in a fiber cable. These principles include Rayleigh scattering and Fresnel reflection theories [(18)]. Several other techniques which are based on the OTDR working principle, such as Tunable-OTDR (T-OTDR), Optical Frequency Domain Reflectometer (OFDR), coherent anti-stokes Raman scattering, has been deployed to in the measures distance of fault in an underground fiber cable [(19)]. There are delays in the use of OTDR measurement alone in the fault tracing process. The delays mostly translate into huge revenue loss to the MONs.

Not so much has been done with machine learning to trace failures in underground optical networks. The quality of transmission in optical systems was estimated by [(20)], by using support vector machine and synthetic datasets. The authors illustrated how LSTM memory could be used for performance prediction after lightpath establishment using field bit error rate data. This process failed to address the fault tracing phenomenon in the underground fiber cable transmissions.

By the use of LSTM RNN based on individual cell profile, [(21)] deployed a generation technique which was able to detect the root cause of faults in an optical system. This technique showed possible strength in detecting both short and long interval abnormal behaviour in the network. The literature of [(22)], analyses a Laser degradation of signal in fiber cable, which is an essential process for the enhancement of laser reliability. In their analysis, [(22)] provided a data-driven fault detection approach based on LSTM recurrent neural networks to detect the different laser degradation modes based on synthetic historical failure data. LSTM model was employed for video transmission throughput prediction. Throughput prediction is essential for ensuring a high quality of service for video streaming [(22,23)]. The processes touched on mainly soft failures and how to detect the associated failures in the fiber cables but refused to address the essential needs in the faults tracing processes. Tracing the exact location of failures in underground fiber network is seldom pursued by researchers in the industry.

In their simulation and presentation, [(24)] designed an intelligent model based on a linear regression technique to predict the actual fiber cut distance on earth. The predictive model had few outliers, which was as a result of the values of the input datasets. In order to overcome this problem identified in [(24)], we leveraged on the cell states in the LSTM deep learning approach to predict the actual distance of fiber cable cut in an underground optical network.

3 Underground optical networks

The demand for internet solutions in the world is continuously rising exponentially, where individuals and agencies depend primarily on this technology for reliable communications. The fastest internet connectivity has been established practically by the use of fiber cable as a transmission medium.

However, to reach as many users possible, the number of fiber cable networks must be expanded worldwide. This expansion will significantly improve service and increased utilization. The expansion of these networks has been achieved by a long and short transmission of the fiber cable. The fiber cable layout for long-distance and short-distance transmission is mostly achieved by an underground transmission system [(25)]. The size, depth and topology of these networks vary by MNOs. Therefore, MNOs are continually seeking to implement innovative techniques for improved, flexible, cost-saving, easy to install, and repair solutions of underground optical networks. As indicated in the simple transmission topology in Figure 1, the ring optical transmission scheme is deployed in underground network infrastructure. The point to point link, as shown in Figure 2 has the optical transmitter plants at node A and the optical receiver plants at node B.

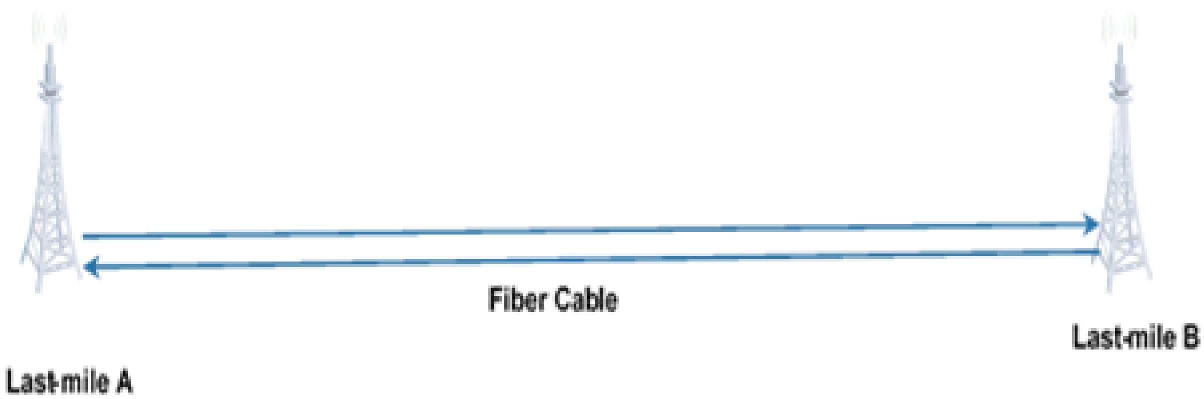

Fig 2. Point-to-point underground optical network

3.1 Machine learning model

An optical network produces a large number of different data streams that must be fetched, processed, and analyzed promptly by machine learning (ML) techniques to ensure QoT [26] in optical network infrastructure. ML is a data analysis technique that automates analytical model building. It is a branch of artificial intelligence based on systems modelling that can learn from available data, identify patterns, and make decisions with minimal or no human intervention [27].

Generally, as long as the underground fiber cable cut remains unsolved, the MNO loses enormous revenue, and users also suffer unnecessarily as a result of the devastating impact of the failure. This paper uses a deep learning model, which is a subfield of the ML technique to predict the actual location of faults in underground optical networks by applying a long-short term memory model.

3.1.1 Deep learning model

The long short-term memory networks (LSTM) networks which is a deep learning model, belong to the family of the recurrent neural network (RNN) model has shown to perform significantly well than other deep learning models. In fact, in one of the most recent and comprehensive books on deep learning, the authors made an emphatic claim that gated RNNs are the most effective sequence models used in practical applications, i.e., LSTMs and gated recurrent unit (GRU) networks [8,9]. There several deep learning predictive models, but we this study adopted an LSTM deep learning model because it is a type of RNN that uses special units in addition to standard units, which makes it easier to memorize past data in memory. LSTM addresses the vanishing gradient problem of RNN. LSTM units include a 'memory cell' that can maintain information in memory for long periods.

A set of gates are used to control information flow in the memory; that is when to output, and when to forget. This architecture allows the model to learn longer-term dependencies. LSTM is suitable to classify, process, and make predictions based on the quality of dataset fed into the model. It trains the model by using back-propagation- when weights are fed back to the input unit to update. Several deep learning models were considered appropriate for this study; however, the LSTM model was adopted due to its ease of use, efficiency, faster training and testing time, which results in low computational cost.

Three gates are present in an LSTM network, that is the input gate, forget gate and output gate.

In the equations (1) - (6), the lowercase variables represent vectors. Matrices w_q and u_q contain, respectively, the weights of the input and recurrent connections, where the subscript q can either be the input gate i, output gate o, the forget gate f and the memory cell c, depending on the activation being computed [28].

Input gate

This LSTM gate decides which value from input should be used to modify the memory [8]. Sigmoid function determines which values to let through 0 or 1, and the tanh function also gives weight to the values which pass, deciding its level of importance ranging from-1 to 1. The relationship between the input gate activation vector and the cell input activation vector, has been given in equation (1) and (2).

$$i_t = \sigma\left(w_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{1}$$

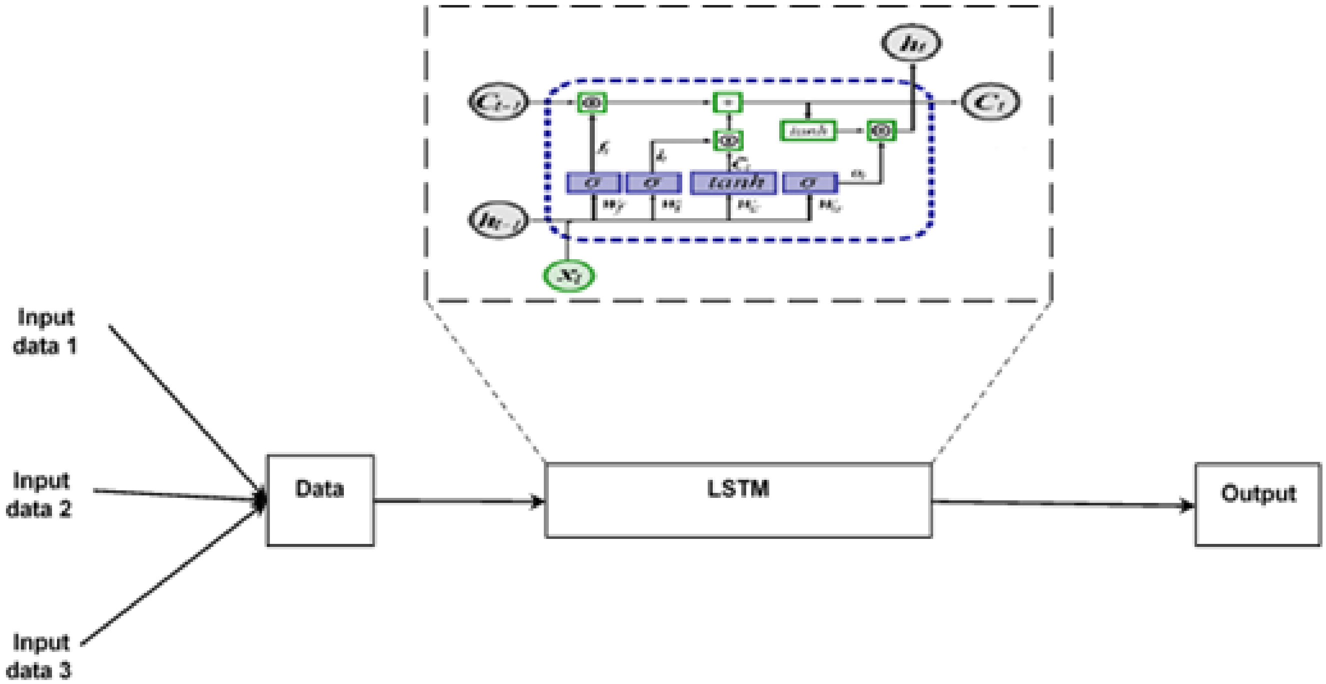

Fig 3. Long short-term memory networks for underground optical network fault prediction

$$\tilde{c}_t = \tanh\left(w_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{2}$$

Forget gate

The forget gate decides what details to be discarded from the block. It is decided by the sigmoid function[10]. It looks at the previous state (h_{t-1}) and the content input (x_t) and outputs a number between 0 and 1 for each number in the cell state c_{t-1}. The mathematical representation of the forget gate activation vector has been given in equationa (3).

$$f_t = \sigma\left(w_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{3}$$

Output gate

The input and the memory of the block of the LSTM model are used to decide the output of the model. Sigmoid function decides which values to let through 0 and 1, and tanh function gives weight to the values which pass to determine its level of importance ranging from -1 to 1 and multiplied by the output of Sigmoid[9] as shown in equation (4), (5) and (6).

$$o_t = \sigma\left(w_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{4}$$

$$t_t = o_t * tanh(c_t) \tag{5}$$

$$c_t = \left(f_t \, o \, c_{t-1} + i_f \, o \, \widetilde{c}_t\right) \tag{6}$$

Where the operator o denotes the Hadamard product and the subscript t indexes the time

f_t is the forget gate activation vector

x_t is the input vector to the LSTM

i_t is the input/update gate activation vector

o_t is the output gate activation vector
h_t is the hidden state vector also known as output vector of the LSTM unit
$\widetilde{c}_t$ is the cell input activation vector
c_t is the cell state vector
σ is the sigmoid function
$w,\ u\ and\ x$ are the weight matrix and bias vector parameters which need to be learned during training

4 Research Design

In this section, we present the methodologies adopted to achieve the objectives of this study. The MNOs in Ghana encounter an average of three thousand (3000) underground fiber cable cuts in every six months. Averagely, two hundred (200) fiber cable cuts are recorded monthly, which mostly affects close to 38% of all telecommunications network infrastructure in the country. We obtained over four thousand nine hundred and three (4903) datasets on previous fiber cable cuts from one of the MNOs in Ghana for this study. The datasets were recorded failures which occurred between Jan. 2015-Dec. 2019. However, after data pre-processing, only four thousand one hundred and eleven (4111) datasets on fiber cable cuts remained. The features of the datasets collected include The stock of cables in initial chamber (β_0), the stock of cables in enclosure box (q), the distance between chambers (α), number of cuts between chambers (N), stock of cable in chamber (K), the measurement of OTDR (x) and the actual distance of failure (Y).

The datasets were pre-processed by using data-cleaning and data-transformation, data quality assessment, feature aggregation, feature sampling, dimensionality reduction, feature encoding techniques. At the data pre-procession stage, rows with missing data were removed, duplications were removed, and feature aggregations were performed in order to put the data in a better perspective. In order for our deep learning model to perform better by reducing computational time, the datasets were scaled to the range of [0, 1] using the maxima-minima function.

The datasets were taken from eight (8) regions in Ghana, where the cuts of the fiber cable were predominantly high. The regions are R1, R2, R3, R4, R5, R6, R7, and R8. The distribution of causes of fiber cable cut in each region has been shown in Table 1.

Table 1. Distribution of cable cuts and the cause of the cut

Cause of fiber cable cut	Cut prevalence in the regions							
	R1	R2	R3	R4	R5	R6	R7	R8
Road Expansion	356	118	319	193	251	59	108	43
Road Construction	201	72	94	92	139	73	67	58
Private developer	64	83	89	72	173	45	79	21
Cable defect	17	9	3	0	2	0	8	11
Water Pipe laying	138	36	103	84	87	31	73	15
Rodent chewing	11	7	2	0	3	0	4	0
Dig Ups	134	49	78	77	53	18	51	37
Railway Construction	22	0	0	0	79	0	0	0

Description of causes underground fiber cable cut

Road expansion: deliberate action which is taken to increase, extend or broaden the existing road

Road construction: the building of roads

Private developer: An individual, a firm, corporation or entity which acquires building or landed property for construction purposes.

Cable defect: this is a fault which is caused by internal damage to an underground fiber cable

Water pipe laying: laying pipes along or across the road

Rodent chewing: when animals deliberately chew part of the underground fiber cable

Dig ups: when electricity power companies dig trenches to mount electricity poles and the road

Railway Construction: When existing track infrastructure receives an expansion, or a new track is under construction.

4.1 Computation of the intelligent model

The diagram in the figure is an illustration of an underground transmission of the optical network. The transmission system comprises a link AB which has plants Transmission (Tx) node and Receiver (Rx) node with six (6) chamber (i.e. C_1, C_3, C_3, C_4, C_5, C_6), equal spacing interval (α_0, α_1, α_2, α_3, α_4, α_5, α_6) between the chambers, splicing enclosure box.

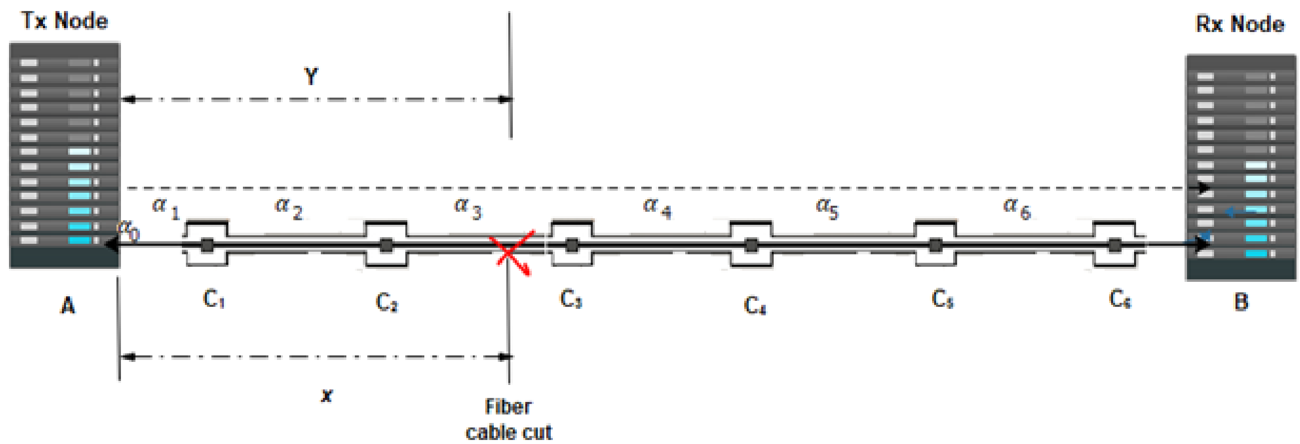

Fig 4. Underground transmission of the optical network

Based on the dataset we used for our LSTM predictive modelling, we developed an algorithm to enable the LSTM network to produce an accurate result. Equations (7) to (14) were is an algorithm developed from the graphical representation of an optical transmission system in Figure 4.

β_0 represents the length of initial coiled cable
α_0 represents the initial distance of the underground fiber cable=0
q_0 represent the stock of cable in enclosure box

N represents the number of chambers between A and B
α_1 represent the length from A to Chamber 1 or from Chamber n to point B
α_2 represent the length from chamber 1 to chamber 2
q_1 represent the stock of cable in enclosure box 1

The length of the cable cut in each interval
$N_0(\mathrm{t})$ represent the number of cuts between A and chamber 1 at a time (t)
$N_1(t)$ represent the number of cuts between chamber 1 and chamber 2 at a time (t)
.
.
.
$N_{n-1}(t)$ represent the number of cuts between chamber n-1 and n
$N_n(t)$ represent the number of cuts between chamber n and B

The length of the coil cable in each chamber
$K_i(t)$ represent the stock of cable in chamber i=1, 2, … , n at a time (t)$K_0 = 0$

$$K_1(t) = \beta_0 - N_0(\mathrm{t}) - N_1(t) \tag{7}$$

$$K_2(t) = \beta_0 - N_1(\mathrm{t}) - N_2(t) \tag{8}$$

$$K_3(t) = \beta_0 - N_2(\mathrm{t}) - N_3(t) \tag{9}$$

.

.

.

$$K_N(t) = \beta_0 - N_{n-1}(\mathrm{t}) - N_n(t) \tag{10}$$

Suppose there is a cut at a point with a distance of x relative to A,

Let, i_x= {1,2,3,…n} be such that

$$\sum_{i-1}^{i_x}\left(K_{i-1}(t) + \alpha_{i-1} + q_{i-1}\right) < x \tag{11}$$

and

$$\sum_{i-1}^{i_x}\left(K_i(t) + \alpha_i + q_i\right) = x \tag{12}$$

Earth distance from A =Y

$$Y = \sum_{i=1}^{i_x-1}\left(\alpha_i + C_i\right) + L_x \tag{13}$$

Where

$$L_x = x - \sum_{i=1}^{i_x-1}\left(K_i(t) + \alpha_i + q_i\right) \tag{14}$$

Where x is the distance of cable measured by the OTDR.

4.2 Application of LSTM model

LSTM predictive model comprises memory blocks called cells, also known by its subdivisions as the cell state and the hidden state was applied to the developed algorithm. The cell state is the main channel of data flow; it allowed the datasets to flow forward through the assigned channel, virtually unchanged. The input datasets were added to or removed from the cell state by sigmoid gates. A gate is similar to a layer or a series of matrix operations, which hold different individual weights[(29)]. The designed LSTM model used gates to regulate the memorizing process. The first step in constructing an LSTM network was to identify information that was not needed; these were subsequently omitted from the cell in that step. These gates determined which datasets entered the long-term memory and then the activation function for the gate. The input gate was a sigmoid function which had a range of (0, 1).

An intelligent predictive model was developed based on the LSTM deep learning technique for this work. The designed LSTM model is excellent for holding long term memories. Preserving the long term dependencies in the LSTM network was achieved by its gating mechanisms. The network stored or released memory on the go through the gating mechanism. As presented in Figure 5, the input dataset was collected and pre-processed, after which a pattern extraction on the dataset was executed before the data was partitioned into training and test dataset. In a deep learning modelling, the more training datasets the model have, the better the model performs, thence, we partitioned our datasets into 80% (3289 datasets) and 20% (822 datasets) of the original dataset, respectively, in order to achieve a better performing model. Data pre-processing is a vital step to obtain better performance and accuracies of machine learning as well as deep learning-based models. Data pre-processing is about dealing with inconsistent, missing, and noisy data. The LSTM predictive model was examined using the mean square error (MSE)[(30)]. We used a sci-kit learn library in Python to model our LSTM network.

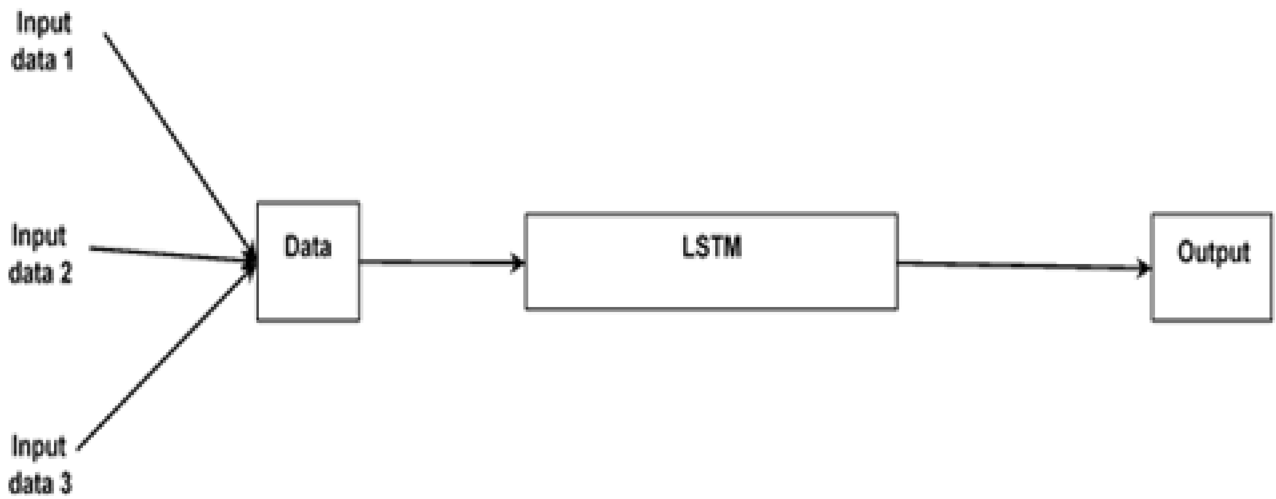

Fig 5. Application of LSTM model to predict the distance of fault

LSTM technique was selected because we found that, LSTM as a recurrent neural network can reminisce values over arbitrary intervals, and it is well-suited to classify processes, and correctly predict the distance of underground fault. It is leveraging on the long-term memory of LSTM, which can allow information from the previous process to be stored in the LSTM memory cells. This is the reason why we adopted the LSTM predictive model because when the model is exposed to new data, it will be able to adapt autonomously. The model learns from the previous computations to produce reliable decisions and produce results for the exact distance of fault on earth. The main activation functions for the LSTM model was Sigmoid and Tanh functions.

4.3 Evaluation Index

The evaluation matrix used to check the correctness and accuracy of the LSTM model was mean squared error (MSE). MSE has been a perfect evaluation matrix for a deep learning model, such as LSTM. We used MSE because of the relations of the dataset we used. MSE is the mean of the squared difference between the predicted value and the actual value. MSE is a risk function, corresponding to the predicted value of the squared error loss. The lower the MSE value, the better it is, and to obtain zero (0) means the LSTM model is perfect[31]. The computed MSE of our LSTM model was achieved by using equation (15).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{15}$$

Where $\widehat{y_i}$ is the predicted distance of fiber cable cut and y_i is the actual distance of fiber cable cut.

5 Result and Discussion

The methods used in this study are based on data-guided approaches and are entirely different from previous studies. Our processes and predictive outcomes will assist the telecommunications industry, especially those that suffer severely from service interruption due to underground fiber cable cut. Figure 6 shows the graphical representation of the actual and the predicted distance. The actual distance is the physical distance of the exact spot of the fiber cable cut as identified by the fiber technicians, y_i . The predicted value also represents the exact physical distance of the place of the fiber cable cut, $\widehat{y_i}$. The y-axis represents the distance (m) of fiber cable cut, and the x-axis represents the number fault that occurred within the specified period of the observation.

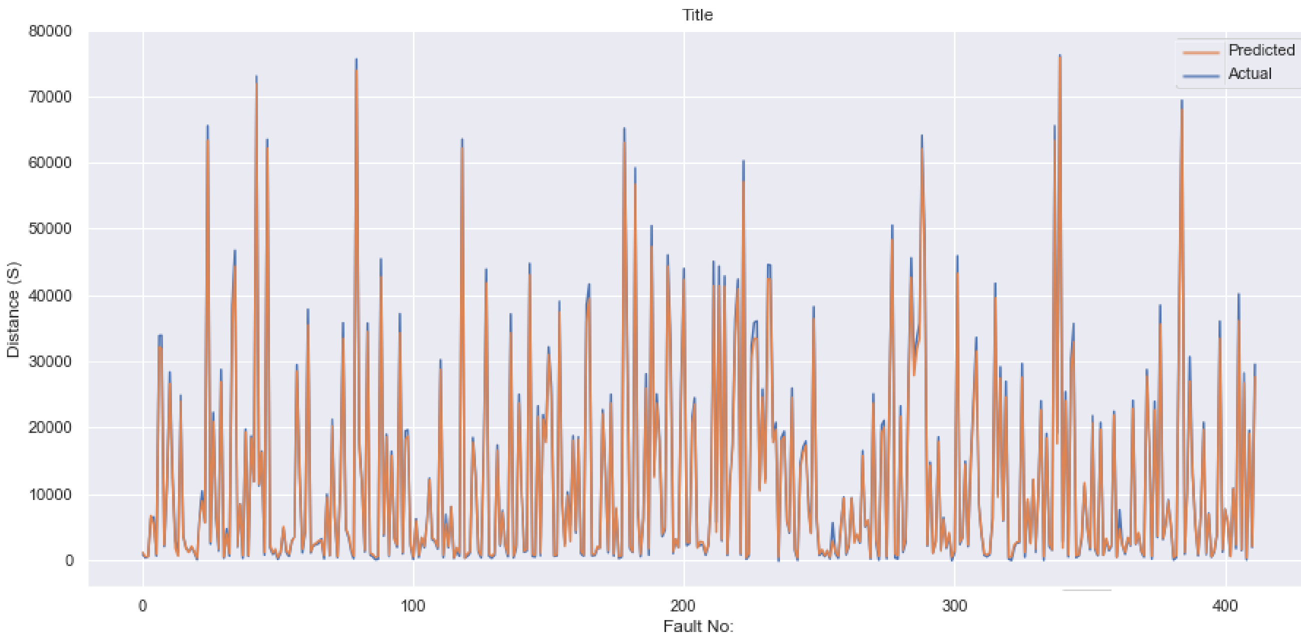

Fig 6. Predicted values of the exact physical distance of the spot of the fiber cable cut

In the LSTM model, the structure of the hidden layer unit is more complicated, usually including multiple active layers, add operations, and multiplication operations. In the simulation of this paper, the LSTM with a different number of layers was used to find the best training model in 100 epochs. During the training process, data truncation length was set to 50, and the parameter distribution for the training are shown in Table 2.

Table 2. LSTM model layers

Layer (type)	Output Shape	Param #
1. lstm_1 (LSTM)	(None, 5, 50)	10400
dropout_1 (Dropout)	(None, 5, 50)	0
2. lstm_2 (LSTM)	(None, 5, 50)	20200
dropout_2 (Dropout)	(None, 5, 50)	0
3. lstm_3 (LSTM)	(None, 5, 50)	20200
dropout_3 (Dropout)	(None, 5, 50)	0
4. lstm_4 (LSTM)	(None, 50)	20200
dropout_4 (Dropout)	(None, 50)	0
5. dense_1 (Dense)	(None, 1)	51

1. Total params: 71,051
2. Trainable params: 71,051
3. Non-trainable params: 0

5.1 Performance matrix

Mean Square Error was used to measure the accuracy of our predictive model for the epoch in the entire process of datasets training. The best ten MSE values for the LSTM model has been presented in Table 3. From the table, the best MSE value was recorded at epoch 89, where the model had its best MSE value.

Table 3. Error function of the LSTM model

epoch	MSE
35/100	1.7279e-04
36/100	1.7331e-04
41/100	1.7841e-04
56/100	1.7698e-04
62/100	1.7838e-04
75/100	1.7857e-04
84/100	1.7116e-04
86/100	1.7059e-04
89/100	1.6673e-04
90/100	1.6676e-04

MSE is a standard and widely used metric for assessing the quality of predictors which is always non-negative, and values closer to zero (0) are better. MSE is the average squared difference between the predicted values and the actual value. This indicates that we achieved the best LSTM error function of 0.00016673 as compared to the results obtained by[24]. The computational time of the LSTM model was better than the linear regression model used by[24]. The MSE value of 0.061291, obtained by[24] indicate that the result obtained in this study is far better than what was obtained when linear regression was used. By applying two intelligent techniques to the same dataset, we achieved varied accuracy report and based on these result we can conclude that LSTM performed better in terms of predicting the exact location of failure in underground fiber networks.

5.2 Correlation matrix

A correlation matrix is a graphical representation showing correlation coefficients between datasets. Each cell in the graph shows the correlation between the two data. A correlation matrix is used to indicate the similarity of datasets, as an input into a more advanced analysis, based on an observable pattern of the dataset.

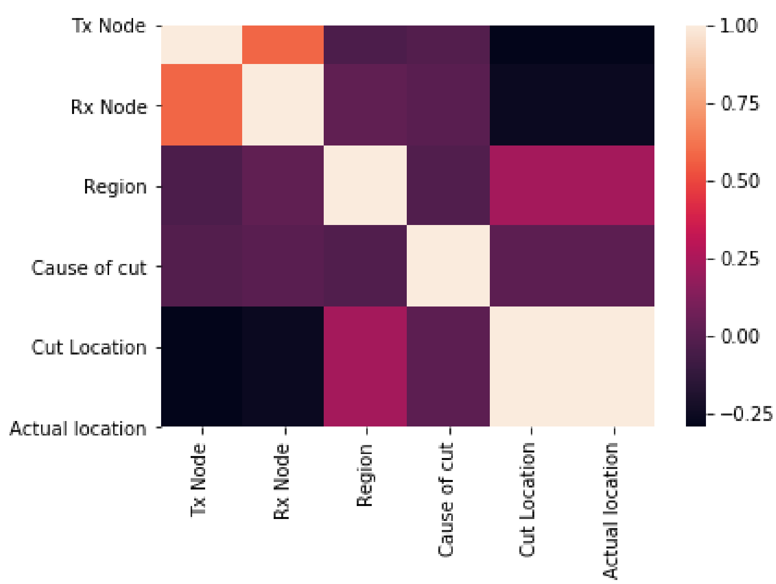

Fig 7. The correlation of the dataset used in the LSTM model.

The six features of the LSTM model have good correlation coefficients, as indicated in Figure 7. The range of values for the correlation coefficient must be between -1.0 to 1.0. That is, the values cannot exceed 1.0, and it not be less than -1.0. However, a correlation of -1.0 indicates a perfect negative correlation, and that of 1.0 indicates a perfect positive correlation.

6 Conclusion

Fault tracing in the underground optical network has been tough over the years, but the outcome of this study has brought a deep relief to the industry player. The adoption of LSTM predictive model has been able to overcome the challenges of using OTDR to trace fault in underground optical networks. Comparatively, the fallout of this paper predicted close the exact distance of fault is better than the fault distance measured by the OTDR. This method has improved the accuracy of tracing fault in an underground optical transmission. The advantage of our approach as compared to the OTDR measurement technique is the additional dataset we used to in the LSTM model to predict the exact location of the fiber cable cut on the earth's surface.

We presented an LSTM process, which made the predicted data more comprehensive and more accurate. The result of this study has given an improved fault tracing process of hard failure in fiber cable. When the work is applied to an active transmission optical networks failures, it has the ability to produce a more precise and accurate prediction of the exact spot in an underground network hard failures.

We have shown that machine learning techniques can bring new insight into the study and prediction of nonlinearity in underground optical transmission systems. Specifically, we have demonstrated that a recurrent neural network with long short-term memory can learn the complex dynamics associated with the fault tracing processes in underground optical networks.

The outcome of the prediction shows that fault in the underground optical networks is predictable with high accuracy. The features of the datasets deployed in this study clearly assume our predictive model is a perfect fault tracing technique than just using the measurement of OTDR. This suggests that the model fits well to the dataset. Moreover, since these metrics measure the degree of closeness between the actual values and the predicted values, based on these results it can be said that the faults in underground optical networks can be predicted accurately with our predictive model. In our future work, we hope to apply gated recurrent unit (GRU) and Graph Convolutional Generative Adversarial Networks (GCN-GAN) non-linear models to our datasets and then compare the results.

References

1) Nyarko-Boateng O, Xedagbui FEB, Adekoya AF, Weyori BA. Fiber optic deployment challenges and their management in a developing country: A tutorial and case study in Ghana. *Engineering Reports*. 2020;2(2). Available from: https://dx.doi.org/10.1002/eng2.12121.
2) Hjort J, Poulsen J. The Arrival of Fast Internet and Employment in Africa. *American Economic Review*. 2019;109(3):1032–1079. Available from: https://dx.doi.org/10.1257/aer.20161385.
3) Thrane J, Wass J, Piels M, Diniz JCM, Jones R, Zibar D. Machine Learning Techniques for Optical Performance Monitoring From Directly Detected PDM-QAM Signals. *Journal of Lightwave Technology*. 2017;35(4):868–875. Available from: https://dx.doi.org/10.1109/jlt.2016.2590989.
4) Etudaye MM, Azi SO. Measurement of signal losses on optical fibre cable due to vibrations using optical time domain reflectometer. *Ghana Journal of Geography*. 2019;11(2):185–198.
5) Hayford-Acquah T, Asante B. Causes of Fiber Cut and the Recommendation to Solve the Problem. *IOSR Journal of Electronics and Communication Engineering*. 2017;12(01):46–64. Available from: https://dx.doi.org/10.9790/2834-1201014664.
6) Ali T, Ameen JH. Study of Fault Detection Techniques for Optical Fibers. *Zanco Journal of Pure and Applied Sciences*. 2019;31(s3):143–149. Available from: https://doi.org/10.21271/ZJPAS.31.s3.20.
7) Zhao Q, Xia L, Wan C, Hu J, Jia T, Gu M, et al. Long-haul and high-resolution optical time domain reflectometry using superconducting nanowire single-photon detectors. *Scientific reports*. 2015;5.
8) Nunes M, Gerding E, Mcgroarty F, Niranjan M. Nunes, Manuel and Gerding, Enrico and McGroarty, Frank and Niranjan, Mahesan, The Memory Advantage of Long Short-Term Memory Networks for Bond Yield Forecasting . 2019. Available from: http://dx.doi.org/10.2139/ssrn.3415219.
9) Goodfellow I, Bengio Y, Courville A. Deep Learning. and others, editor;MIT Press. 2016.
10) Eriksson TA, Bulow H, Leven A. Applying Neural Networks in Optical Communication Systems: Possible Pitfalls. *IEEE Photonics Technology Letters*. 2017;29(23):2091–2094. Available from: https://dx.doi.org/10.1109/lpt.2017.2755663.
11) Argyris A, Bueno J, Fischer I. Photonic machine learning implementation for signal recovery in optical communications. *Scientific Reports*. 2018;8(1). Available from: https://dx.doi.org/10.1038/s41598-018-26927-y.
12) Kushnir D, Gohil G, Sayeed Z, Uzunalioglu H. Predicting outages in radio networks with alarm data. *IEEE/ACM International Symposium on Quality of Service*. 2018.
13) Zhu M, Ye K, Xu CZ. Network anomaly detection and identification based on deep learning methods. In: and others, editor. CLOUD. 2018;p. 219–234.
14) Nakamura A, Okamoto K, Koshikiya Y, Manabe T. Potential Fault Detection in Optical Cables Using OTDR Operating in Two-Modes. In: and others, editor. 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC). 2019. Available from: https://doi.org/10.23919/ps.2019.8817759.
15) Wu J, Lee PC, Li Q, Pan L, Zhang J. CellPAD: Detecting Performance Anomalies in Cellular Networks via Regression Analysis. In: Proceedings of IFIP Networking. 2018.
16) Bensalem M, Singh SK, Jukan A. Machine Learning Techniques to Detecting and Preventing Jamming Attacks in Optical Networks. 2019.
17) Ilouno J, Awoji M, Kwaha B, Chagok N. Comparative Analysis on Fault Detection Techniques in Fiber Optics Communication Links. 2018.

18) Ali TA, Ameen JJH. Study of Fault Detection Techniques for Optical Fibers. *ZANCO Journal of Pure and Applied Sciences*. 2019;31(s3):143–149.
19) Xia TJ, Wellbrock GA, Huang M, Salemi M, Chen Y, Wang T, et al. First Proof That Geographic Location on Deployed Fiber Cable Can Be Determined by Using OTDR Distance Based on Distributed Fiber Optical Sensing Technology. *Optical Fiber Communication Conference (OFC) 2020, OSA Technical Digest*. 2020.
20) Tremblay C, Allogba S, Aladin S. Quality of transmission estimation and performance prediction of lightpaths using machine learning. In: 45th European Conference on Optical Communication (ECOC). 2019.
21) Mamun A, Beyaz SMA, M. LSTM Recurrent Neural Network (RNN) for Anomaly Detection in Cellular Mobile Networks. *Lecture Notes in Computer Science*. 2019;p. 222–237. Available from: https://doi.org/10.1007/978-3-030-19945-6_15.
22) Abdelli K, Rafique D, Pachnicke S. Machine Learning Based Laser Failure Mode Detection. *21st International Conference on Transparent Optical Networks (ICTON)*. 2019. Available from: https://doi.org/10.1109/icton.2019.8840267.
23) Wei B, Kawakami W, Kanai K, Katto J, Wang S. TRUST: A TCP Throughput Prediction Method in Mobile Networks. In: 2018 IEEE Global Communications Conference (GLOBECOM). 2018;p. 1–6. Available from: https://doi.org/10.1109/GLOCOM.2018.8647390.
24) Nyarko-Boateng O, Adekoya AF, Weyori BA. Predicting the actual location of faults in underground optical networks using linear regression. *Engineering Reports*. 2020.
25) Salmela L, Tsipinakis N, Foi A, Billet C, Dudley JM, Genty G. Predicting ultrafast non-linear dynamics in fibre optics with a recurrent neural network. 2020.
26) Samadi P, Amar D, Lepers C, Lourdiane M, Bergman K. Quality of transmission prediction with machine learning for dynamic operation of optical WDM networks. *2017 European Conference on Optical Communication (ECOC)*. 2017;p. 1–3.
27) Maimo LF, Gomez ALP, Clemente FJG, Perez MG, Perez GM. A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks. *IEEE Access*. 2018;6:7700–7712. Available from: https://dx.doi.org/10.1109/access.2018.2803446.
28) Wang C, Fu S, Xiao Z, Tang M, Liu D. Long Short-Term Memory Neural Network (LSTM-NN) Enabled Accurate Optical Signal-to-Noise Ratio (OSNR) Monitoring. *Journal of Lightwave Technology*. 2019;37(16):4140–4146. Available from: https://dx.doi.org/10.1109/jlt.2019.2904263.
29) Qin X, Yang C, Zhou Q, Tian F, Feng J, Xu C, et al. Low-complexity Bi-directional Recurrent Neural Network Equalizer for Short-Range Optical Interconnect Links. *Asia Communications and Photonics Conference*. 2019;p. 4–70.
30) Wang D, Zhang M, Zhang Z, Li J, Gao H, Zhang F, et al. Machine Learning-Based Multifunctional Optical Spectrum Analysis Technique. *IEEE Access*. 2019;7:19726–19737.
31) Aladin S, Allogba S, Tran AVS, Tremblay C. Recurrent Neural Networks for Short-Term Forecast of Lightpath Performance. *Optical Fiber Communication Conference*. 2020;p. 2–24.