

CASE STUDY

Open Access



Using machine learning techniques to predict the cost of repairing hard failures in underground fiber optics networks

Owusu Nyarko-Boateng* , Adebayo Felix Adekoya and Benjamin Asubam Weyori

*Correspondence:
owusu.
nyarko-boateng@uenr.edu.gh
Department of Computer
Science & Informatics,
University of Energy
and Natural Resources,
Sunyani, Ghana

Abstract

Fiber optics cable has been adopted by telecommunication companies worldwide as the primary medium of transmission. The cable is steadily replacing long-haul microwave, copper cable, and satellite transmissions systems. Fiber cable has been deployed in an underground, submarine, and aerial architecture to transmit high-speed signals in intercontinental, inter countries, inter cities and intra-cities. Underground fiber cable transmission has experienced major failures as compared to other mediums of fiber transmission infrastructure. The failure is rampant, and especially the cable get cuts frequently in areas where there are road constructions, road road expansion projects, and other developmental projects. The cost of repairing these failures is enormous, and it largely depends on the cause of failure and the geographical area the faults occurred. The main aim of this paper was to investigate the cost of repairing underground fiber cable failures, clustered the cause of faults, and then used feedforward neural networks (FFNN) and linear regression to predict the cost of repairing future faults. The result of the predictive model is significant to the telecommunications industry, which means the cost of repairing an underground optical networks will be known to the industry players before the fault occurs. depending on which area, the cause of the failure and the mean time to repair (MTTR), the predictive model tells the mobile network operators the cost involved to repair the damaged cable. The accuracy of the result obtained indicates the predictive model is good for predicting the cost of repairing fiber cable cut in underground optical networks.

Keywords: Cost of repairing fiber cable, Underground optical networks, Cause of cable cut, Feedforward neural networks, Linear regression, MTTR, MTBF, K-means clustering

Introduction

The two most common outdoor fiber optic cable installations are aerial installation and underground cable installation. Underground cable installation is buried directly underground or placed into a buried duct. Underground burial of fiber cable installations is mostly common for long-distance installations. The cables are usually buried in trenches. Underground duct installation also provides an opportunity for future expansion without the need to dig again. Preparation towards underground cable

installation requires the adherence to certain outlined guidelines which includes obtaining proper right-of-way permits, Identify existing underground utilities such as buried cables, pipes, Investigate the soil condition to determine the installation depth, whether duct should be used and the type of fiber cable to be used [1].

The cost of deploying underground fiber cable is extremely expensive. The depth of the underground fiber cable installation is between 1 to 2 m. The soil conditions and its type determines the depth the fiber cables should be buried. In colder areas, fiber cables are typically buried below the frost line to prevent the cables from being damaged by ground frost heaves. Through forests, rivers, drylands, etc. fiber cable has been laid in a seemingly protective manner. These protections are usually not sufficient to prevent the cut of the underground cable [2].

Previous research works have shown that fiber cables were more likely to be cut in rural areas, even though urban areas have a higher number of cables cuts [1, 2]. Cutting the underground fibre optic cables causes service interruption leading to a considerable inconvenience to businesses and home users. Underground fiber cable cuts reduce network reliability, affects customer experience on the network. The cut also increases the service provider's operational costs which further reduces rural network expansion, affects last mile connectivity strategies, and it also incurs a high maintenance cost [2]. The cost of maintenance or repairing cuts in an underground cable is very expensive [3]. This paper sought to determine the cost of repairing fiber cable cuts by using K-means clustering for failure classification and FFNN to predict the cost of repairing fiber cable in the future. Knowing the cost of repairing a faulty underground cable by the industry players helps them to budget and plan ahead of failures.

First, we considered the Mean Time Between Failures (MTBF) of the underground fiber transmission links. We assume this basic unit to be the minimum length between the transmitter and the receiver. Values for MTBF are hard to obtain from literature.

Fiber networks are repairable systems with a mean time to repair (MTTR). The values of the MTTR depend on a variety of factors such as duct type (aerial, underground or submarine cable), failure location (city area or rural/hard-to-reach areas), kind of the damage and, of course, reaction and completion times defined in the contracts with the responsible repair companies. In the instances stated the use of MTTR values in the range of 1 h (Best case) to 4 h (worst case) [4]. Assume both MTBF and MTTR to be the same throughout the network. Thus a fiber transmission link unit has $\text{availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$.

Case description

The failure is rampant, and especially the cable gets cuts frequently in an area where there is road construction work, the road expansion project, and other developmental projects. The cost of repairing these failures is enormous, which prompted this paper to investigate the impact of underground fiber cable failures, cluster the cause of faults, and then used feedforward neural networks (FFNN) and linear regression to predict the cost of repairing the faults. The predicted value will inform the industry players to know the cost involved in repairing any faults outlined in this paper.

Literature review

The cost prediction this work entails using machine learning and deep learning models, for instance, neural networks and K-means to obtain the expected result. ML algorithms receive and analyze input data to predict output values. They improve their performance while being fed with new data, collect and analyze high volumes of data about fiber cable cuts. Deploying of ML modelling is a new research area, so no literature has so far been identified concerning the topic presented in this paper. However, there were few works of literature which provided a cost-effective predictive model in big data process and the quality of result [5]. Ahmadvand and Goudarzi [6], also presented their simulation a quick and effective mechanism to rank portions of big data processing steps for cost-effective purposes. The warehouse-scale computers used more time- and cost-efficiently process data, or process partial data with optimized approximation concerning resource, time and energy limitations. Ahmadvand and Goudarzi [6] showed that different data portions, from the same or different sources, have different significance in determining the outcome of the computation, and hence, by prioritizing them and assigning more resources to the processing of more essential data. A cost-effective predictive model using adapting genetic algorithm was adopted by Padhy, Singh and Satpathy [7] to predict the fault tolerance in web service applications.

In an optical network transmission system, the cost of installation and deployment of cables and plants is quite expensive. However, critical protective measures have been adopted over the years to safeguard the various component in the optical transmission ecosystem from damage. That notwithstanding, negligence on the part of some organizations and developers have lead to the destruction of many optical cables. When cables are destroyed by known or unidentified developers, the affected mobile network operator (MNO) has to fund the repairing of the optical cable quickly. This situation has lead to unimaginable loss of funds to the MNO, which arguably increases the Capex of the MNO. The optical cables are mostly buried underground for a long-haul transmission system. When a cut occurs in the underground cable, the technician team are required to dig up the cable and repair it. The cost of repairing a fault in an underground optical cable is unknown to the MNOs. The use of machine learning to predict the cost of repairing a fault in fiber optic cable has not attracted the attention of the scientific research community. This study uses the ML model to predict the cost of repairing cuts in underground optical cables.

Hard failures in underground optical networks

Despite the massive infrastructure and the meshed optical networks built to support the deployment of the services, mobile network operators (MNOs) continue to suffer fiber cable cuts, which disrupts service quality and interrupts service to users. Hard failures are referred to as the physical damage to the underground fiber optics cable such as cable breaks or cuts due to several activities [1, 2]. The telecommunications industry in Ghana records a monthly average of 200 underground fiber cable cuts, which affect close to 38% of all network interruptions leading to the obstruction of quality of service targets annually. The cuts of underground fiber cables reduce the optical network availability, affect customer experience, and increases the MNOs

cable repair costs. The rampant underground cable cut further reduces rural network expansion and affects last mile connectivity strategies [1–3].

Private developers and road construction are the leading cause of the destruction of underground fiber cables. Additionally, drainages, dig-ups and other utility service organizations which access the road reservation corridors have contributed immensely to the underground fiber cable destruction. It is crucial to note that continuous cuts in underground fiber cable result in degradation of the optical signal as a result of the additional splicing joints due to constant repairs. This phenomenon severely affects the quality of service (QoS). In ensuring QoS, MNOs are confronted with the challenge of replacing long spans of fiber cables to eliminate the multiple splicing joints, and this comes at huge costs and extended time in restoring network outages [2].

Other developing nations such as Nigeria, Kenya, India, etc. have experienced similar hard failures in the underground optical networks. Studies have found that underground fiber cables were more likely to be cut in rural areas, even though several unapproved activities in urban areas have accounted to a higher number of fiber optic cables destructions in a given area [1, 3].

K-mean clustering

K-means algorithm is an iterative algorithm that attempts to partition the dataset into K pre-defined distinct non-overlapping subgroups called clusters, where each data point belongs to only one group. The algorithm cluster data points as similar as possible while also keeping the clusters as different as possible. The algorithm then allocates data points to a cluster in such a way that the summation of the squared distance between the data points and the cluster's centroid. The less variation obtained within clusters, the more homogeneous the data points are within the same cluster. The clustering technique is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. In this paper, the K-means algorithm deployed grouped the dataset into C_n clusters [8].

Let C_1, \dots, C_k denote sets containing the indices of the dataset in each cluster. These set satisfying two properties, according to [9]:

$C_1 \cup C_2 \cup \dots \cup C_k = \{1, \dots, n\}$, each data belong to one of the K clusters.

If the i th data is in the K th cluster, then $i \in C_k$ cluster C_k is a measure $w(C_k)$ of the amount by which the data within a cluster differ from each other as derived by [10] in Eqs. (1), (2), (3) and (4) ie:

$$\text{minimize}_{C_1, \dots, C_k} \left\{ \sum_{k=1}^k w(C_k) \right\} \quad (1)$$

Squared Euclidean distance will be of the clusters is:

$$w(C_k) = \frac{1}{|C_k|} \sum_{i, j \in C_k} \sum_{j=1}^p (x_{ij} - x_{i,j})^2 \quad (2)$$

where $|C_k|$ denotes the number of the dataset in the K th cluster

$$\text{minimize}_{C_1, \dots, C_k} \left\{ \sum_{k=1}^k \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'kj})^2 \right\} \quad (3)$$

There are K^n ways to partition n datasets

$$\frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'kj})^2 = 2 \frac{1}{|C_k|} \sum_{i \in C_k} \sum_{j=1}^p (x_{ij} - \bar{x}_{kj})^2 \quad (4)$$

where

$$\bar{x}_{kj} = \frac{1}{|C_k|} \sum_{i \in C_k} x_{ij}$$

is the mean for feature j in the cluster C_k , which minimizes the sum-of-squared deviation. However, the coefficient of deviation or separation took the values in the interval $[-1, 1]$.

If it is 0, then the sample is very close to the neighbouring clusters.

If it is 1, then the sample is far away from the neighbouring clusters.

If it is -1 , then the sample is assigned to the wrong clusters.

Silhouette analysis was used to determine the degree of separation between the clusters. That means the distance from all data points is in the same cluster (k). When the average distance from all data points in the closest cluster (C_k) [10, 11].

Hence, the computation of the coefficient has been given in Eq. (5) as:

$$\text{coefficient} = \frac{k - C_k}{\max(C_k, k)} \quad (5)$$

Feedforward neural networks

Feedforward neural networks fundamentally consist of three layers, which are the input layer, hidden layer, and output layer. The FFNN model has one hidden layer with monotonically increasing differentiable functions, which has the ability to approximate the continuous function with the hidden layer [12]. The FFNN model used in this paper has two input neurons and a hidden layer with two nodes, which comprise of the data collected from the field, which are the determinants of the cost of fiber cable repairs. The input variables include the cause of a fault and the area/region it occurred [13].

The output of the FFNN for the input pattern Z_p was computed with a multilayer forward pass through the network. According to [14] each output unit, O_k , is given in Eqs. (6), (7) and (8) as:

$$O_{k,p} = f_{ok}(net_{ok,p}) \quad (6)$$

$$= f_{ok} \left(\sum_{j=1}^{J+1} w_{kj} f_y (net_{yj,p}) \right) \quad (7)$$

$$= f_{ok} \left(\sum_{j=1}^{J+1} w_{kj} f_{y_j}, \left(\sum_{i=1}^{I+1} v_{ji} z_{ip} \right) \right) \quad (8)$$

where f_{ok} and f_{y_i} are the activation functions, O_k is the output unit, y_i is the hidden unit, W_{kj} is the weight between the output unit O_k and the hidden unit y_j , $Z_{i,p}$ is the value of the input unit z_i of the input pattern z_p .

The $(I+1)$ th input unit and the $(J+1)$ th hidden unit were the bias units representing the threshold values of the neurons in the next layer within the hidden layer [15].

The machine learning model designed for our predictive model has been shown in the figure. After applying the transform function to Eqs. 6, 7 and 8, our predictive model now has x_1 , x_2 , and x_3 as the input variables of the mode with $\sum_{i=1}^n (x_i w_i)$ being the activation function [16], as shown in Fig. 1.

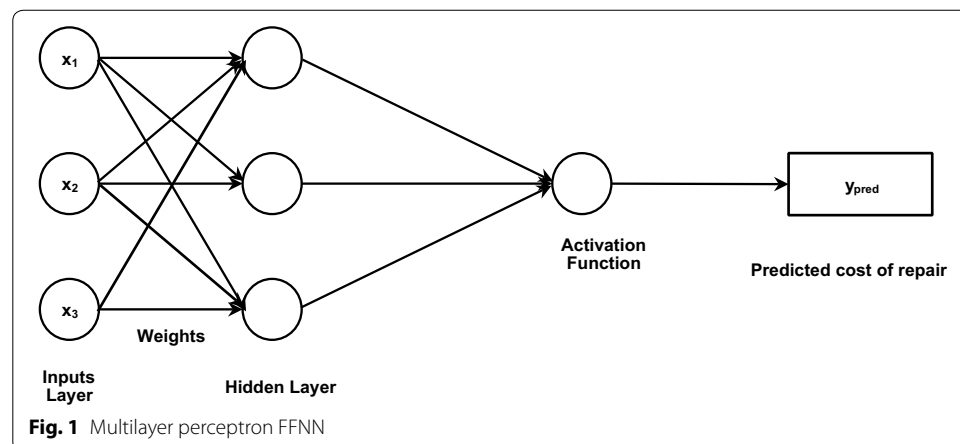
$$y_{pred} = \sum_{i=1}^n (x_i w_i) \quad (9)$$

Thus

$$y_{pred} = \sum_{i=1}^n (x_i w_i) \quad (10)$$

Research design

According to Ghana Chamber of Telecoms [1], the leading MNO in the country experience 2000 underground fiber cable cuts in every 6 months and other MNOs suffers about 1500 cuts over the same period. Averagely, 200 cable cuts are recorded monthly, which mostly affects close to 38% of all network infrastructure. The study was designed to collect the necessary data for simulation, testing and evaluation. Weka 3.8 machine learning simulation tool was deployed. Weka is a collection of machine learning algorithms for solving real-world data mining and big data problems. We used Weka for our simulation because it is an open-source machine learning software with a good graphical user interface, standard terminal applications and contains built-in tools for standard



machine learning tasks, and additionally gives transparent access to well-known tool-boxes such as sci-kit learn. In all, 4111 datasets on fiber cable cuts were collected from the eight regions in Ghana, where the cuts of the fiber cable are 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. The datasets were collected over 4 years, thus, between January 2015 and December 2019. The research was designed.

Description of causes of fiber cable cut

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

Road construction: A project that involves the creation of a new road and drainage system.

Private developers: Any individual, firm, corporation or entity, other than a nonprofit corporation, limited profit entity, or public corporation who acquire buildings or land in order to construct or refurbish building projects on the site.

Cable defect: This is a fault which is caused by internal damage of 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.

Table 2 presents the distribution of the cause of fiber cable cut against the total fiber cable cuts throughout the observation period and the cost of repairing each fault. According to the referenced table, the cost of repairing a fiber cable cut caused by a road expansion, road construction, private developer and so on is \$3800, \$3800, 4200, 4400 respectively. Table 3 shows the total fiber cable cut in each region throughout the observation period and the total amount spent on repairing those faults. Table 3 revealed that region 1 (R1) recorded the highest cable cuts followed closely by region 5 (R5), with the least cable cuts occurring in region 8 (R8). The pattern of the cause of cable cuts and its associated cost of repairs were critically examined and modelled to give an accurate cost value by using machine learning techniques such as k-means clustering, FFNN and

Table 1 Distribution of regional 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

Table 2 Distribution of cause of cable cut against the cost of repairing each cable cut

Cause of cut	Total cuts	Cost of repair (\$)
Road expansion	1447	3800
Road construction	796	3800
Private developer	626	4200
Cable defect	50	4400
Water pipe laying	567	3000
Rodent chewing	27	4400
Dig ups	497	2900
Railway const. works	101	3000

Table 3 Distribution of regional cable cuts and the total cost of repairing cuts in each region

Region	R1	R2	R3	R4	R5	R6	R7	R8
Fiber cable cut	943	374	688	518	787	226	390	185
Cost of repairs	3,583,400	1,421,200	2,614,400	1,968,400	2,990,600	858,800	1,482,000	703,000

linear regression. By simple accounting and mathematical computations, the cost of repair can be made known to the MNOs. However, the use of machine learning modeling provides a better result quickly with preciseness and exactness.

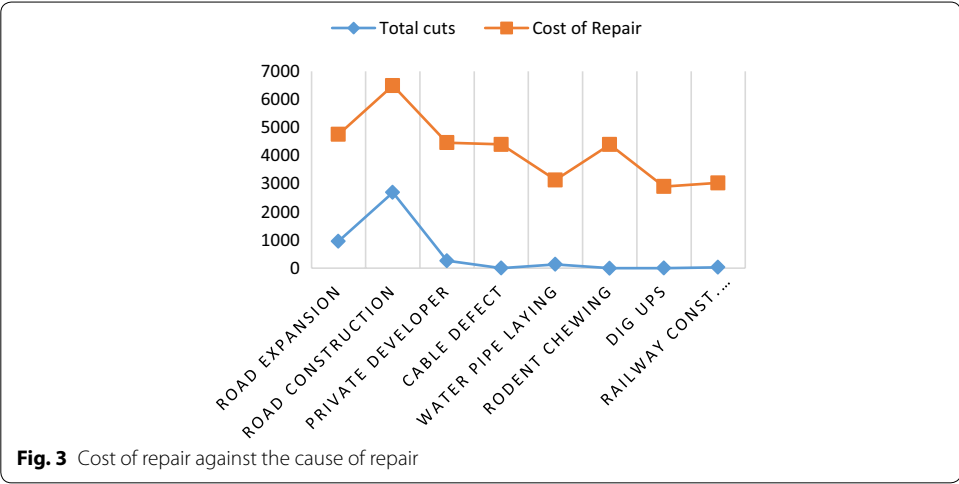
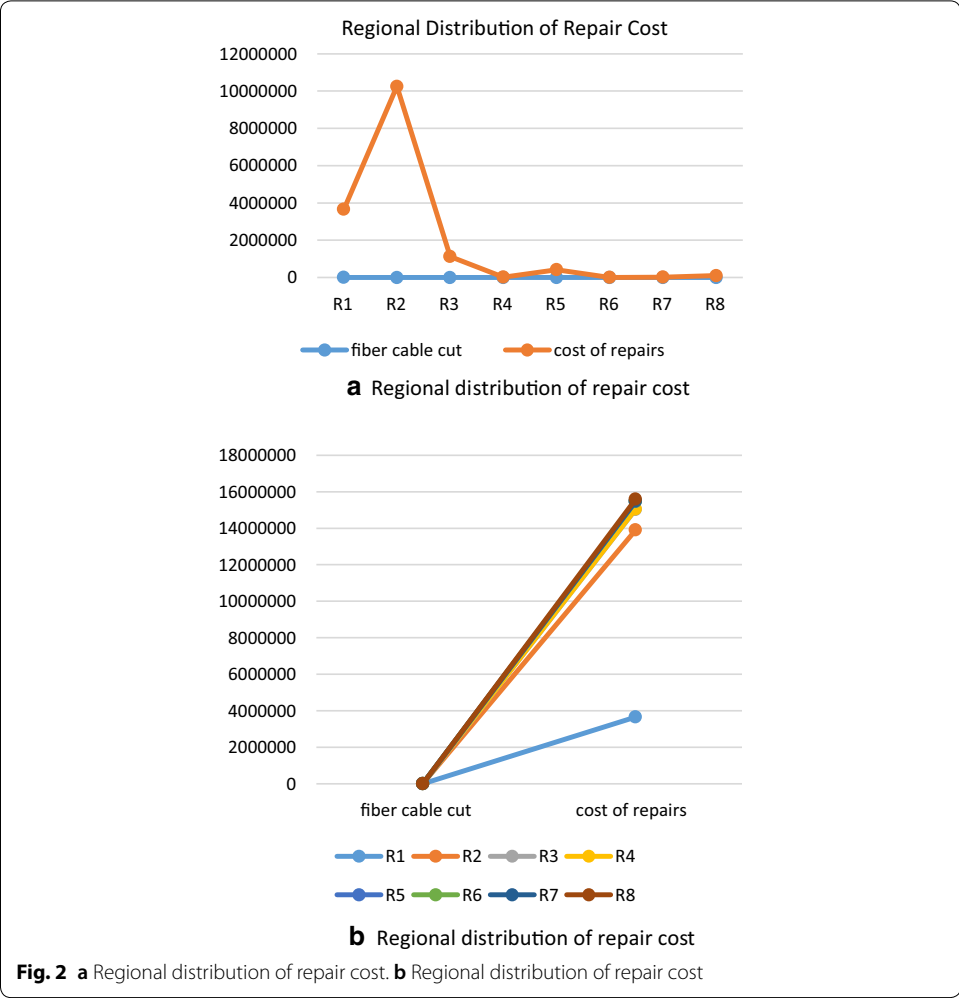
The cost of repairing faulty underground fiber cable largely depends on the cause of cut, the location, the MTTR, and the texture of soil where the cable has been buried. The location could be dryland, across the road and waterloo. The soil texture is sandy, rocky and clayey. All these factors directly affect the cost of repairs and even may cause delays in tracing the fault [17, 18].

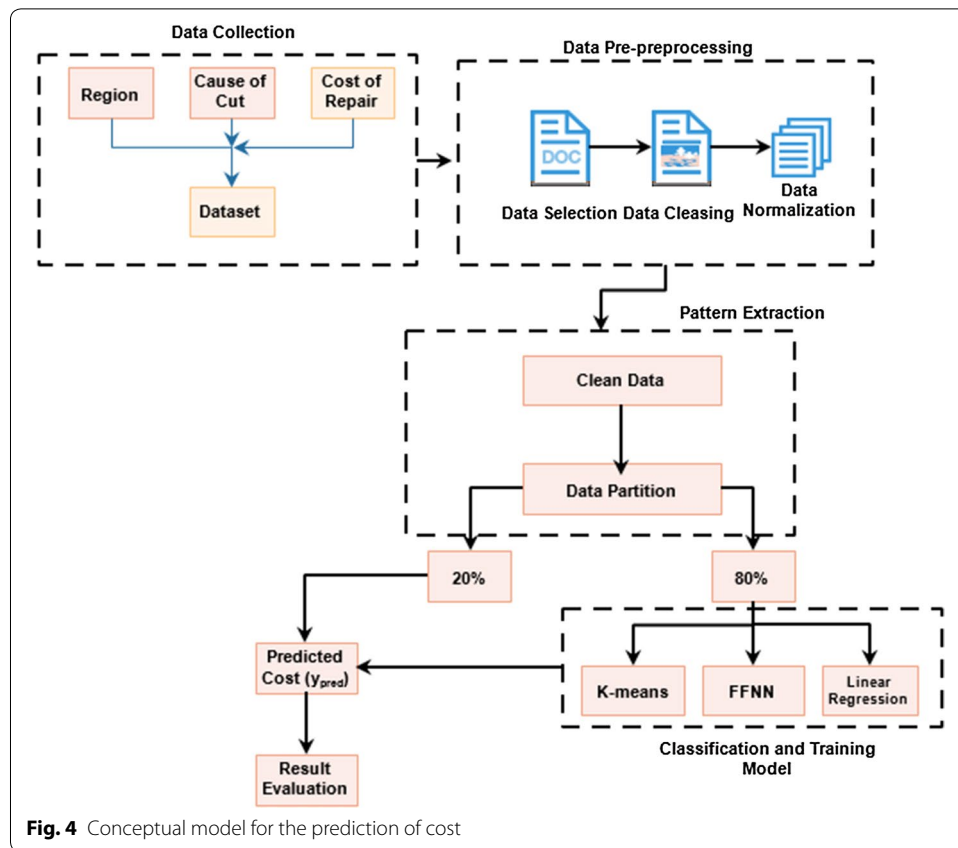
The graphical representation of the predictive model has shown Fig. 2a, b. The cost of repairing a faulty cable largely depends on the cause of cable cut and at which location the event occurred. The higher the MTTR, the larger the cost of repair of the fault. In Fig. 2a, it has been indicated that the highest cost of cable repair was in R2, which was a result of a considerable number of cable cuts within the period of observation [18].

However, Fig. 3 shows the comparison of the cost of repair against the total cables for each cause of the cut. It has been established in Fig. 3 that road construction was the highest cause of cable cut which also accounted for the total highest cost of repairing cable cuts.

Conceptual model for the prediction of cost

The paper collected data from the field in eight regions. The field datasets included the cause of fiber cable cuts, the distance of cuts, time of cuts, etc. The cost of repairing a fiber cut according to the Ghana Chamber of Telecoms is \$3000. The datasets were integrated and pre-processed by data selection, data cleansing, and data normalization. The cleaned dataset was partitioned into 80% training dataset and 20% test dataset, as shown in Fig. 4. K-means clustering, FFNN and linear regression were the machine learning algorithms [19] with sigmoid activation in the Weka library were applied to the datasets





to classify and predict the cost of repairing cut cables. The iterative characteristic of an ML model is very significant in the sense that as the ML model is exposed to new datasets, they can adapt autonomously. The model quickly learns from the previous computations of the cost of repairing faults and cause of the cable cut to produce reliable and high accurate results. We adopted the ML technique to simulate our work because of its ability to classify, integrate and train new dataset into the predictive model to produce the expected output. The ML model was carefully chosen to aid in providing good accuracy in our cost prediction research.

The cost of repairing fiber cable cut (y) = the cost of repair + the revenue lost * MTTR

$$Availability = \frac{MTBF}{MTBF + MTTR} \quad (11)$$

MTBF is the average time elapsed from one failure to the next [4].

MTTR is the average time taken to repair faults after a failure. In Ghana, the MTTR for fault repair is ≤ 4 h.

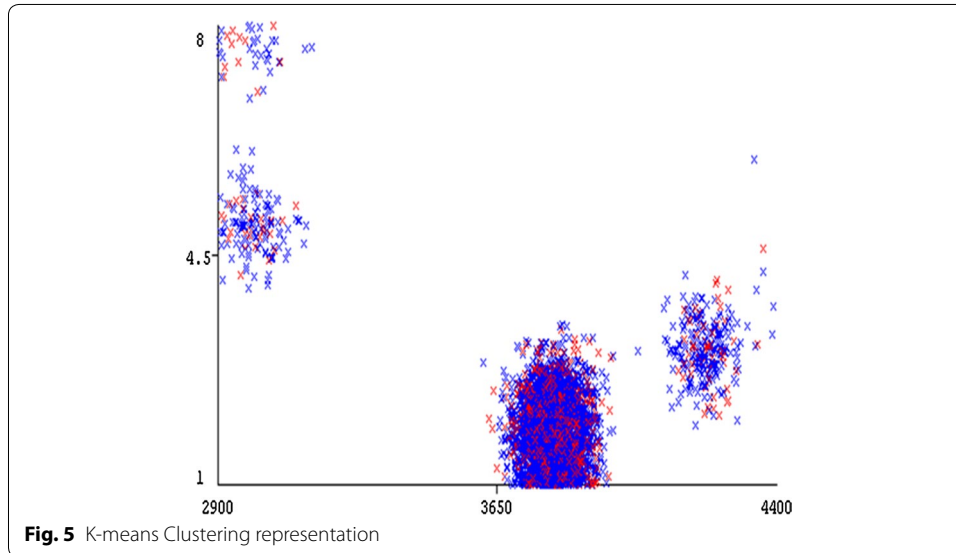
Calculating actual MTBF requires a set of observations of fiber cable cuts; each observation is:

Uptime: The moment at which an optical network began operating (initially or after a repair).

Downtime: The moment at which an optical network fail operating as a result of cut since the previous uptime-moment.

Table 4 Cluster centroid

Attribute	Full data (4111.0)	0 (3165.0)	1 (946.0)
Region	3.3126	2.3169	6.6438
Cause of cut	1.9927	1.994	1.9884
Cost of cut repairs	3791.8755	3791.4692	3793.2347



So each Time Between Failure (TBF) is the difference between one Uptime_moment observation and the subsequent Downtime_moment.

Three quantities are required:

n = number of observations. u_i = This is the i th Uptime_moment. d_i = This is the i th Downtime_moment following the i th Uptime_moment.

So Mean Time Between Failures = $\sum \frac{d_i - u_i}{n}$, for all $i = 1$ through n cable cuts. More simply, it is the total working time divided by the number of failures.

Discussion and evaluation

The K-means clustering technique applied to the dataset produced final cluster centroids are shown in clustering Table 4 and Fig. 3. Table 4 present two main cluster in the clustering process; Cluster 0 and Cluster 1, respectively. The clusters have the same attributes which have been represented by Region, Cause of cut and Cost of cut repair. Figure 5 shows the graphical representation of the various clustering and the degree of association. Silhouette analysis was used to determine the degree of separation between the clusters and the distance from all data points in the same cluster.

The Classifier model (full training set) produced the result which made the clustering of the fiber cable cut, causes of cut and the cost of repairs a goon one. Tables 5, 6, 7 and 8 represent the nodes of the neural networks used in the predictive model and the threshold for the weight and inputs, respectively. The result obtained when the sigmoid

Table 5 Linear Node 0

Inputs	Weights
Threshold	0.1938372813438577
Node 1	− 7.623539323585559
Node 2	6.561431557746884
Node 3	3.324653334879235

Table 6 Sigmoid Node 1

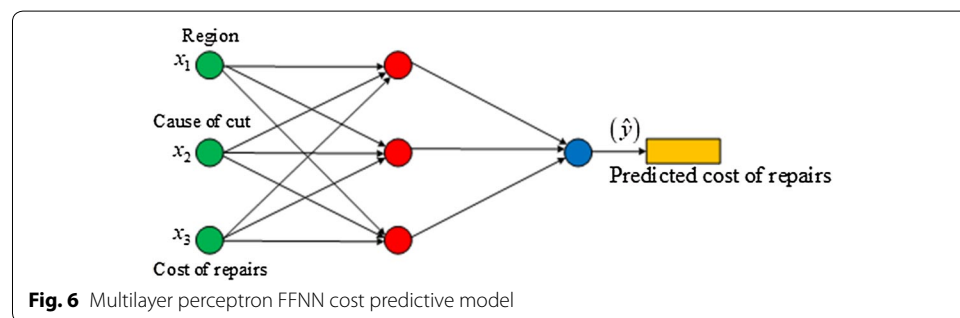
Inputs	Weights
Threshold	3.5412435862303884
Attrib region	− 0.011259471397618341
Attrib cause of cut	19.860635797639173

Table 7 Sigmoid Node 2

Inputs	Weights
Threshold	4.443522517572205
Attrib region	− 0.003043989816696887
Attrib cause of cut	15.760480062916114

Table 8 Sigmoid Node 3

Inputs	Weights
Threshold	4.5517422862303114
Attrib region	− 0.021243571398713722
Attrib cause of cut	16.320635545634532



activation function was applied to each node in relation to the weight of the inputs variables indicate a good correlation in the features deployed in the predictive model.

A linear regression model [20, 21] was used to train the multilayer perceptron FFNN that produced the predicted cost of repairing a cut cable, as shown in Fig. 6. Three inputs were fed into the multilayer perceptron FFNN cost predictive model, which was computed using Eq. (9). The output of the model [22–24] was the predicted cost of repair, which accurately

correlate with the pattern of the clustered input data. The predicted cost [25–27] of repairing an underground fiber cable cut, according to the model, can be determined when variables such as region and cause of cuts are known. The empirical machine learning model [28] we used in this paper requires an update of recent fault variables to be able to predict the actual cost of repairing the cable.

Results evaluation

Two main evaluation metrics were adopted to measure the performance of the predictive model. The Mean Absolute Error (MAE) and Root mean squared error (RMSE) were the two metrics used to measure the accuracy of the predicted cost of repairing fiber cable cut. The accuracy rate obtained was perfect for the predictive model. MAE measured the average magnitude of the errors in a set of predictions while RMSE is a quadratic scoring rule which was deployed to measure the average size of the error.

Both MAE and RMSE expressed average model prediction error which means metrics ranged from 0 to ∞ and are indifferent to the direction of errors. The metrics are negatively-oriented scores, which means lower values are better. Equations (12), (13) and (14) were used in the computation of the error matrices [10, 11].

$$MEA = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (13)$$

$$CC = \frac{n(\sum y_j \hat{y}_j - (\sum y_j)(\sum \hat{y}_j))}{\left[\sqrt{\sum y_j^2 - (\sum y_j)^2} \right] \left[\sqrt{n \sum \hat{y}_j^2 - (\sum \hat{y}_j)^2} \right]} \quad (14)$$

The predictive model has been able to predict the cost of repairing an underground fiber cable cut with a high accuracy rate. The MAE value obtained, as indicated in Table 9 shows an accurate prediction as this paper sort to achieve. We achieved the best value of 0.0658 for MAE performance matrix. The obtained MAE value show that our model has high accuracy in predicting the cost of repairing underground fiber cables. The result attained is extremely significant in the telecommunication industry, which experiences frequent underground fiber cable cut to determine the cost of repairing the cable as quickly as possible. The telecommunication company will be able to allocate the exact predicted cost of repair for the maintenance of the cable. Using a machine learning model for such industrial practice offers significant relief of unknow budget and unreliable overhead cost, which may stifle the MNOs Opex.

Table 9 Evaluation of the prediction model

CC	MAE	RMSE
0.5179	0.0658	8.8574

The correlation coefficient (CC) indicates the strength of the relationship between the actual cost of repairing fiber cable and the predicted cost of repairing a fiber cable.

Additionally, the accuracy rate obtained from the result was perfect for the predictive model. The two main evaluation metrics used to measure the accuracy of the prediction were the Mean Absolute Error (MAE), and Root means squared error (RMSE). According to [10] and [11], lower values of MAE and RMSE are better. This assertion strongly supports the results of the model deployed in this paper which gave the value of the correlation coefficient as 0.5179, indicating that the features of the dataset used the prediction were highly correlated. The performance metric deployed met the empirical threshold of measurement, and that makes our result more relevant to predict the expected cost when hard failures occur in underground optical network infrastructure.

Conclusion

The main aim of this paper was to investigate the cost of repairing underground fiber cable failures, classify the cause of faults, and then used the ML model to predict the cost of repairing future faults. We achieved this carefully collecting dataset and then deployed the Weka ML simulation tool to achieve our objectives. The result of the predictive model is significant to the telecommunications industry, which means the cost of repairing an underground optical network will be known to the industry players before the fault occurs. Depending on which area, the cause of the failure and the MTTR, the predictive model, tells the mobile network operators the cost involved to repair the damaged cable.

Road construction, private developers, water pipes installation, railway construction, etc. are some of the reasons for underground fiber cable cuts as indicated in the dataset used for predictive modelling. The K-means clustering model was able to provide balanced clustering with accurate centroid computation. The multilayer perceptron FFNN was analyzed and evaluated with the best MAE value, which made the result of this paper have good, acceptable accuracy in predicting the cost of repairing fiber cable cuts. The accuracy of this work shows the MNOs will now have an idea about the cost of repairing the cut in an underground optical cable before it happens. The cost of repairs depend on the features used for prediction, thus, the region and the cause of the cut.

There has been series and rampant fiber cable cuts in Ghana, parts of Africa and other low-income countries. The obtained MAE value show that our model has high accuracy in predicting the cost of repairing underground fiber cables. The result attained is extremely significant in the telecommunication industry, which experiences frequent underground fiber cable cut to determine the cost of repairing the cable as quickly as possible. The telecommunication company will be able to allocate the exact predicted cost of repair for the maintenance of the cable. Using a machine learning model for such industrial practice offers significant relief of unknow budget and unreliable overhead cost, which may stifle the MNOs Opex.

Abbreviations

FFNN: Feedforward neural networks; NN: Neural networks; MTTR: Mean time to repair; MTBF: Mean Time Between Failure; QoS: Quality of service; MNO: Mobile network operator; TBF: Time between failure; CC: Correlation coefficient; MAE: Mean absolute error; RMSE: Root mean square error.

Acknowledgements

We thank the leading Telecommunications company in Ghana, which gave us the data.

Authors' contributions

ONB was involved in data collection, predictive model designing, and data pre-processing. ONB, AFA, and BAW were major contributors in writing and approving the manuscript. All authors read and approved the final manuscript.

Funding

No funding was obtained for this paper.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available because the mobile network operator who released the data do not want it to go public but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests" in this section.

Received: 19 April 2020 Accepted: 14 August 2020

Published online: 24 August 2020

References

1. AllAfrica. Ghana: Fiber cuts—root of poor service quality and negative user experience. 2020. <https://allafrica.com/stories/201902040673.html>. Accessed 14th Feb 2020.
2. Mata J, de Miguel I, Durán RJ, Merayo N, Singh SK, Jukan A, Chamania M. Artificial intelligence (AI) methods in optical networks: a comprehensive survey; 1573–4277. *Opt Switch Netw*. 2018;28:43–57.
3. 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. *Eng Rep*. 2020;2(2):e12121.
4. Alavian P, Eun Y, Liu K, Meerkov SM, Zhang L. The (α, β) -Precise estimates of MTBF and MTTR: definitions, calculations, and induced effect on machine efficiency evaluation. *IFAC-PapersOnLine*. 2019;52(13):1004–9. <https://doi.org/10.1016/j.ifacol.2019.11.326>.
5. Ahmadvand H, Goudarzi M. SAIR: significance-aware approach to improve QoR of big data processing in case of budget constraint. *J Supercomput*. 2019;75(9):5760–81.
6. Ahmadvand H, Goudarzi M. Using data variety for efficient progressive big data processing in warehouse-scale computers. *IEEE Comput Archit Lett*. 2017;16:166–9.
7. Padhy N, Singh RP, Satapathy SC. Cost-effective and fault-resilient reusability prediction model by using adaptive genetic algorithm based neural network for web-of-service applications. *Clust Comput*. 2018;9:1–23.
8. Ahn E, Kumar A, Feng D, Fulham M, Kim J. Unsupervised feature learning with K-means and an ensemble of deep convolutional neural networks for medical image classification. 2019. arXiv preprint [arXiv:1906.03359](https://arxiv.org/abs/1906.03359).
9. Zöller MA, Huber MF. Survey on automated machine learning. 2019. arXiv preprint [arXiv:1904.12054](https://arxiv.org/abs/1904.12054).
10. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer Science+Business Media; 2013. <https://doi.org/10.1007/978-1-4614-7138-7>. ISBN 978-1-4614-7138-7.
11. Ouellet S, Michaud F. Enhanced automated body feature extraction from a 2D image using anthropomorphic measures for silhouette analysis. *Expert Syst Appl*. 2018;91:270–6. <https://doi.org/10.1016/j.eswa.2017.09.006>.
12. Orimoloye LO, Sung M-C, Ma T, Johnson JEV. Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Syst Appl*. 2020;139:112828. <https://doi.org/10.1016/j.eswa.2019.112828>.
13. Panayiotou T, Chatzis SP, Ellinas G. Leveraging statistical machine learning to address failure localization in optical networks. *IEEE/OSA J Opt Commun Netw*. 2018;10(3):162–73.
14. Engelbrecht AP. Computational intelligence: an introduction. Hoboken: Wiley. 2007. SN - 9780470512500. <https://books.google.com.gh/books?id=IzoslGJMjUC>
15. Mäkinen M, Iosifidis A, Gabbouj M, Kannianen J. Predicting jump arrivals in stock prices using neural networks with limit order book data. 2018. SSRN 3165408.
16. Ghobadi M, Mahajan R. Optical layer failures in a large backbone. In: IMC '16: proceedings of the 2016 internet measurement conference November 2016. 2016. pp. 461–7. <https://doi.org/10.1145/2987443.2987483>
17. Tran DT, Magris M, Kannianen J, Gabbouj M, Iosifidis A. Tensor representation in high-frequency financial data for price change prediction. In: 2017 IEEE symposium series on computational intelligence (SSCI). IEEE. 2017. pp. 1–7.
18. Ntakaris A, Mirone G, Kannianen J, Gabbouj M, Iosifidis A. Feature engineering for mid-price prediction forecasting with deep learning. 2019. arXiv preprint [arXiv:1904.05384](https://arxiv.org/abs/1904.05384).
19. Christensen MH, Nozal MH, Kavadas I, Pinson P. Data-driven learning from dynamic pricing data—classification and forecasting. In: 2019 IEEE Milan PowerTech, Milan, Italy, 2019. pp. 1–6. <https://doi.org/10.1109/PTC.2019.8810769>
20. Wang Di, Guo X, Guan C, Li S, Jinhui Xu. Estimating stochastic linear combination of non-linear regressions efficiently and scalably. *Neurocomputing*. 2020. <https://doi.org/10.1016/j.neucom.2020.02.074>.
21. Funke B, Hirukawa M. Bias correction for local linear regression estimation using asymmetric kernels via the skewing method. *Econometr Stat*. 2020. <https://doi.org/10.1016/j.ecosta.2020.01.004>.
22. Anitha P, Patil MM. RFM model for customer purchase behavior using the K-Means algorithm. *J King Saud Univ Comput Inf Sci*. 2019. <https://doi.org/10.1016/j.jksuci.2019.12.011>.

23. Rahouma KH, Ali A. Applying machine learning technology to optimize the operational cost of the Egyptian optical network. In: 16th international learning & technology conference 2019. *Procedia Comput Sci.* 2019;163:502–17. <https://doi.org/10.1016/j.procs.2019.12.133>. 1877–0509 © 2019 The Authors. Published by Elsevier B.V.
24. Hong Y, Hammad AW, Akbarnezhad A. Forecasting the net costs to organizations of Building Information Modelling (BIM) implementation at different levels of development (LOD). *J Inf Technol Constr.* 2019;24(33):588–603.
25. Utatunda S. Machine learning: an introduction. In: Laha A, editor. *Advances in analytics and applications.*, Springer Proceedings in Business and Economics; Singapore: Springer; 2019. https://doi.org/10.1007/978-981-13-1208-3_1.
26. Nousi P, Tsantekidis A, Passalis N, Ntakaris A, Kannianen J, Tefas A, Gabboujlosifidis MA. Machine learning for forecasting mid-price movements using limit order book data. *IEEE Access.* 2019;7:64722–36.
27. Ntakaris A, Magris M, Kannianen J, Gabbouj M, Iosifidis A. Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. *J Forecast.* 2018;37(8):852–66.
28. Tsantekidis A, Passalis N, Tefas A, Kannianen J, Gabbouj M, Iosifidis A. Using deep learning for price prediction by exploiting stationary limit order book features. 2018. arXiv preprint [arXiv:1810.0](https://arxiv.org/abs/1810.0)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
