JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

A Three Index Profile Core Dataset for Designing Few Modes Fiber for Long Haul Optical Communications

S. P. Obaje, Z. M. Abdullahi, A. D. Usman, I. Yau, Agbon E. Ezeikel, A. S. Yaro
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria

ABSTRACT

Fiber is widely recognized as the medium for next-generation communication, offering secure transmission and efficient switching capabilities. Few-Mode Fiber (FMF), which accommodates multiple modes with weak coupling, is particularly attractive for Mode Division Multiplexing (MDM) systems. Nevertheless, the multi-parameter characteristics of FMF make its design highly complex, often inaccurate, and time-intensive due to its intricate structure and the presence of numerous higher-order modes. This research work presents a 3-index profile core dataset for FMF design with the aid of Machine Learning (ML) regression model for long haul optical communications applications. The generated dataset for the FMF design parameters was developed for long haul optical fiber length, which span 100 Km to 1000 Km. The FMF parameters used for generating the dataset consists of the effective refractive index of core, the radian distance of the 3-index core profile and the difference between the refractive indices of the 3-index core and that of the cladding. With additional design properties which consist of the fiber birefringence of the propagating modes, the coupling correlation lengths of the optical length, and the coupling coefficient of the propagating modes, the Differential Mode Group Delay (DMGD) of the fiber design and the IM-XT for the five guides modes for the FMF. The realized FMF design dataset was generated aid of numerical equations that characterize FMF parameters and MATLAB software. The dataset of the FMF design parameters consists of 10000 rows and 53 columns with step size of 1×10-4. The simulation analysis carried out for the 3-index FMF design dataset validation shows that, the correlation matrix analysis, for very high positive entries of 10.5%, for high positive entries of 12%, moderate positive entries of 15%, moderate negative entries of 6%, high negative entries of 3%, very high negative entries of 1% and for weak/no entries of 52.5%. Furthermore, the boxplot analysis for the percentage outlier and normal data for each design parameters was presented. They are, overall average percentage outlier of 2.80% and that of normal data of 97.20% was realized. Finally, the percentage cumulative variance for the first Principal Component (PC1) of 63.9795%, that of second the Principal Component (PC2) of 83.2868% and that of third the Principal Component (PC3) is 13.2516% respectively. With the aid of the developed dataset, machine learning models can be trained to predict different optimal design parameters for FMF, therefor minimizing design complexities and enhance precision in FMF designs.

ARTICLE INFO
Article History
Received: September, 2025
Received in revised form: October, 2025
Accepted: January, 2026
Published online: January, 2026

KEYWORDS
FMF Design, Dataset, 3-Index Profile, Correlation Matrix, Boxplot, PCA, long-Haul and Optical Communications

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

INTRODUCTION

Optical fiber is widely known as the medium for next-generation communication, owing to its high-speed, secure transmission, and efficient switching capability (Shreya, 2023). Mode Division Multiplexing (MDM) and Space Division Multiplexing (SDM) offer excellent switching performance and enable significantly higher data transmission rates (Mizuno & Miyamoto, 2017). Among various fiber types, Few-Mode Fiber (FMF), a special class of Multi-Mode Fiber (MMF) that supports multiple weakly coupled modes and is well-suited for long-distance transmission is highly desirable in MDM systems (Pierre et al., 2014). Moreover, FMF can substantially increase the channel capacity of an optical medium (Qiuyan *et al*., 2018), as illustrated in Figure 1.

Despite its advantages, the design of FMF is inherently challenging due to its multi-parameter nature. Complex fiber structures and the presence of numerous higher-order modes make the design process difficult, often imprecise, and time-consuming (Xinyi *et al*., 2022). One of the most critical performance-limiting factors in FMF is Inter-Mode Crosstalk (IM-XT), also referred to as intra-core crosstalk, which arises from coupling among propagating modes within the fiber core (Liu & Xiang, 2024). IM-XT is primarily induced by fiber imperfections, including nonlinear effects, bends, twists, and refractive index perturbations (Luca, 2025). Its severity depends on the number of transmitted modes in the core (Rademacher *et al*., 2022). When only the first three modes (LP01, LP11a, and LP11b) are transmitted, the resulting crosstalk remains relatively low. However, with five-mode transmission (LP01, LP11a, LP11b, LP21a, and LP21b) or higher, the accumulated modal crosstalk increases significantly along the transmission link (Qiuyan *et al*., 2018).

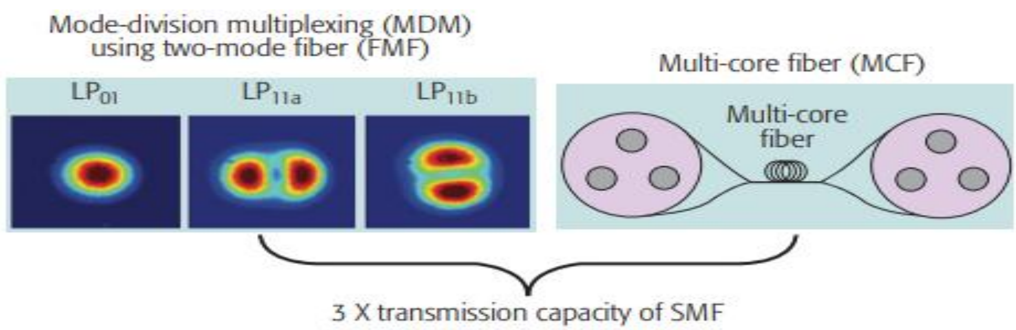

Figure 1: Multi-Modes Transmission in a Few Mode Fiber (Ken-ichi & Nikolaos-Panteleimon, 2017)

Similar Works

The multi-parameter characteristics of FMF make its design highly complex, often inaccurate, time-intensive due to its intricate structure and the presence of numerous higher-order modes which may increase the presence of IM-XT among propagating neighboring modes therefore, compromising the channel capacity. This subsection presents similar works that had been done by researcher to address it.

Zhiqin *et al*., (2020) propose an inverse design method based on Neural Network (NN) to optimize the structural parameters of ring-assisted Few Modes Fiber (FMF) that can accommodate 4, 6, 10 and 20 modes. This was realized by generating a dataset with weak coupling and effective index difference between adjacent modes (Δdesign n_{eff}) greater than 1×10^{-3}. By training the Nural Network (NN), the structural parameters of the optical fiber are calculated instantaneously according to the distribution of the effective refractive index values. This method provides high-accuracy, high-efficiency and with high complexity but fast and reusable design of optical fibers. Bhagyalaxmi *et al*., (2023) presented an inverse modeling technique using regression-based Machine Learning (ML) to design a weakly coupled Few Mode Fiber (FMF) for enabling Mode Division Multiplexing (MDM). The dataset used for training the ML model was

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

generated with the aid of COMSOL simulation software. The parameters of the proposed ring-core FMF are varied within an appropriate range to obtain a specific number of modes with weak coupling between the few adjacent modes.

The parameters (r1, r2, r3, Δ1, Δ2, Δ3) are the input to the Finite Element Method (FEM) based on commercially available software COMSOL to obtain the mode solutions for each guided mode. The dataset that characterized the mode solutions are arranged in a 5000 × 33 matrix for training and testing the ML model. One set of data took 30 minutes for generation and compilation of the dataset. The dataset is generated by varying the range of the r_i and Δ_i design parameters. The range of the design parameters are varied to achieve a maximum of 26 number of modes over an extensive range of n_{eff}. The range of design parameters are selected to achieve a large number of modes and more than 50% of the mode solutions in the dataset satisfies the criteria of the weak mode coupling However, the technique deployed in the work takes a lot of time to obtain the mode solutions for the FMF design.

Junling *et al*., (2024) worked on Panda Polarization-Maintaining Few-Mode Optical Fiber (PPMFMOF) for short distance optical transmission. The dataset used in the work was designed to accommodates 10Lp modes with the aid of forward design method for the design of PPMFMOF based on Artificial Neural Network (ANN). In the study, different ANN models on the fiber performance were examined based on the minimum effective refractive index difference between adjacent LP modes. Finally design of PPMFMOF supporting the 10 LP modes in C + L band was successfully realized. The error mechanism deployed in the work is maximum Mean Absolute Percentage Error (MAPE) to validate the ANN model. The method provided high-efficiency and high-accuracy for the fast design of PPMFMOF though ANN are complex network and high memory consumption.

From the attempts made by previous works, the techniques deploy were able to address design complexity and inaccuracy in the optical design. However, the dataset used for training the ML model does not capture the secondary design parameters that captures impairments that can affect the channel capacity of the FMF. This work presents the generation of a three-index profile core dataset for FMF design, for training ML model for long haul optical communications applications.

Dataset for the FMF Design Parameters

The method deployed to actualized the dataset for the FMF design are discussed in this section. The FMF design parameters from the range of values that satisfy the three-index profile design ranges from the parameter's values of r_1, r_2, r_3, Δ_1, Δ_2 and Δ_3 respectively (Bhagyalaxmi *et al.,* 2023). The are used to generate the dataset in this work. The values for the range of parameters deployed for the development in this work are shown in Table 1. As shown in Table 1, r_1, r_2 and r_3 are the radian distance of the three-index profile. While Δ_1, Δ_2 and Δ_3 are the refractive index difference between the three-index core profile and the cladding (Bhagyalaxmi *et al.,* 2023). With the aid of this initial parameters value the refractive index of the three-ring core can be determined. The parameter values in Table 1 were use as the initial values to obtain other parameters that resulted in generating the dataset of 10000 rows and 53 columns with step size of $1*10^{-4}$ using MATLAB simulation software.

Table 1: Range of Parameters for the three Ring Core Fiber (Bhagyalaxmi *et al.,* 2023)

Parameters	Minimum	Maximum
r_1 (μm)	1.5	6
r_2 (μm)	7.5	12
r_3 (μm)	11	15
Δ_1	0.001	0.026
Δ_2	0.007	0.033
Δ_3	0.002	0.02

Modified Mathematical Model for the Refractive Indices of the 3-Index Core

The refractive index of the 3-index profile core was obtained using a modified mathematical model as shown in Equation (1). The model was obtained by modifying the existing model which support the propagation of 5 to 20

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

modes which result to complex simulation algorithm and design structure in the work of Bhagyalaxmi *et al.,* (2023). While the modified mathematical model is for propagating modes (Lp01, Lp11, Lp21, Lp02 and Lp31) and maximum of 6 modes when the polarized mode for Lp11 (Lp11a and Lp11b) are consider. The modified mathematical mode will reduce complex simulation algorithm and design structure for the five propagating modes in the FMF design.

$$n_{coIM}(r) = \begin{cases} \mathrm{n}_{1IM} = \mathrm{n}_{clad} + \Delta_{1IM} & 0 \le r \le r_{1IM} \\ \mathrm{n}_{2IM} = \mathrm{n}_{clad} + \Delta_{2IM} & \mathrm{r}_{1IM} \le r \le r_{2IM} \\ \mathrm{n}_{3IM} = \mathrm{n}_{clad} + \Delta_{3IM} & \mathrm{r}_{2IM} \le r \le r_{3IM} \end{cases} \qquad (1)$$

where n_{coIM} is the refractive index of the ring-core of the for the improved design, n_{1IM} denotes the refractive index of the first ring for the improved design, n_{2IM} corresponds to the refractive index of the second ring for the improved design, n_{3IM} implies the refractive index of the third ring for the improved design, n_{clad} expresses the refractive index of the cladding for the improved design, r_{1IM} coincides with the radius of the first ring for the improved design, r_{2IM} is the radius of the second ring for the improved design, r_{3IM} represents the radius of the third ring for the improved design, r implies the radius of the core of the FMF, Δ_{1IM} corresponds to the difference between $\boldsymbol{n_{1IM}}$and $\boldsymbol{n_{clad}}$, Δ_{2IM} represents the difference between $\boldsymbol{n_{2IM}}$ and $\boldsymbol{n_{clad}}$, Δ_{3IM} denotes the difference between $\boldsymbol{n_{3IM}}$ and $\boldsymbol{n_{clad}}$.

Cross-Sectional View of the Three-Index Profile Design Parameters

The cross-sectional view of the three-index profile design parameters is presented. The refractive index profile of the 3-index core design as a function of radian distances is presented in Figure 2. The core of the FMF comprises of refractive index of n_{1IM}, n_{2IM} and n_{3IM} while, that of the cladding is n_{clad}. The refractive index of the second core (n_{2IM}) is higher than that of the first core (n_{1IM}). While that of the first core (n_{1IM},) is higher than that of the third core (n_{3IM}). The variation in the refractive index in the fiber core determine the numbers of ring in the FMF core.

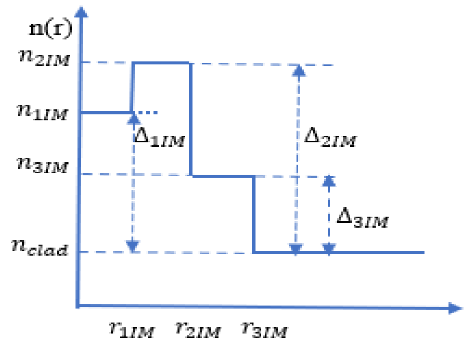

Figure 2: Refractive Index Profile of the 3 Index Core Design

The parameters of FMF were designed to allow a specific number of modes (5 modes) to propagates in the core of the optical FMF. The optimized core parameters will minimize the Inter-Mode Cross-Talk (IM-XT) among the adjacent modes propagating in the FMF. This will enhance the channel capacity of the FMF. The cross-sectional view of the 3-index core profile is shown in Figure 3.

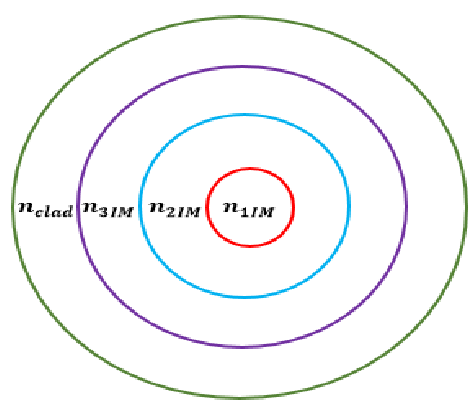

Figure 3: Cross Sectional View of the Three Ring Core Profile of the FMF Design

In Figure 3. n_{1IM}, n_{2IM} and n_{3IM} are the refractive index of the 3-index ring profile of the FMF core while n_{clad} is the refractive index of the FMF cladding. n_{1IM}, n_{2IM} and n_{3IM} are design to have different refractive indices with n_{2IM} the highest, followed by n_{1IM} and n_{3IM} the least

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

among the 3-ringed core. Its helps in containing the propagating modes in each of the ring with related refractive indices.

Developing the Dataset for the three Index Profile Core

The parameters used for developing the dataset for the three-index profile core for the FMF consists of the effective refractive index, the birefringence, the coupling correlation lengths, and the coupling coefficient to determine the Differential Mode Group Delay (DMGD) and the IM-XT for the five guides modes in the FMF are presented.

Effective Refractive Index for the Five Guides Modes

The effective refractive index of the five propagating modes in the FMF core are determined by Equation (2). Equation (2) was obtained from Equation (1) it shows the refractive index, in which the five modes will propagate effectively in the core of the FMF (Bhagyalaxmi *et al.,* 2023).

$$Eff_{\mathrm{ni}} = n_{\mathrm{cladding}} + \Delta_i \qquad (2)$$

Where Eff_{ni} coincides with the effective refractive index of the five propagating modes, Δ_i is the deference in refractive index of the five propagating modes and that of the cladding. n_{cladding} represents the refractive index of the cladding.

The effective refractive index difference of the propagating modes in the 3-index core are determined by subtracting the effective refractive index of the i^th^ propagating mode from that of the j^th^ propagating mode. The absolute values of the difference in the effective refractive index difference are shown by Equation (3).

$$\Delta n_{eff} = |n_{effi} - n_{effj}| \qquad (3)$$

where Δn_{eff} denotes the difference between the effective refractive index of the adjacent mode. n_{effi} represent the effective refractive index of the i^{th} propagating mode, n_{effj} implies the effective refractive index of the j^{th} propagating mode.

Birefringence of the FMF Design

The birefringence of the propagating modes expresses the bipolarization of the propagating mode in the FMF design as expressed by Equation (4) (Arun *et al*., 2019).

$$\text{birefringence} = \max Eff_{\mathrm{ni}} - \min Eff_{\mathrm{ni}} \qquad (4)$$

where $\max Eff_{\mathrm{ni}}$ is the maximum effective refractive index of the i^{th} propagating modes and $\min Eff_{\mathrm{ni}}$ denotes the minimum effective refractive index of the i^{th} propagating modes.

Mode Coupling Correlation Length for Each Neighboring Modes

The mode coupling correlation length for each neighboring modes are shown in Equation (5) (Agrawal G. P., 2010).

$$\mathrm{Lc_{ij}} = \frac{\lambda}{|Eff_{\mathrm{ni}} - Eff_{\mathrm{nj}}|} \qquad (5)$$

where $\mathrm{Lc_{ij}}$ denotes the coupling correlation length for each neighboring mode, λ represents the wavelength of the propagating modes, $|Eff_{\mathrm{ni}} - Eff_{\mathrm{nj}}|$ means the absolute value of the difference between the i^{th} propagating modes and that of the j^{th} propagating neighboring modes.

Coupling Coefficient for Each Neighboring Modes

Mode coupling coefficient of the propagating modes are expressed by Equation (6) (Agrawal G. ,2010).

$$\mathrm{kappa_{ij}} = \frac{|Eff_{\mathrm{ni}} - Eff_{\mathrm{nj}}|}{\lambda} \qquad (6)$$

where $\mathrm{kappa_{ij}}$ represent the coupling coefficient of the i^{th} and j^{th} neighboring modes, λ denotes the wavelength of the propagating modes. $|Eff_{\mathrm{ni}} - Eff_{\mathrm{nj}}|$ means the absolute value of the difference between the i^{th} propagating modes and that of the j^{th} propagating neighboring modes.

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Differential Mode Group Delay for Each Neighboring Modes

The effective group indices of the five propagating modes are expressed by Equation (7) (Linqi *et al.,* 2022).

$$\mathrm{n_{g,i}} = Eff_{\mathrm{g,i}} + \lambda \times 0.1 \qquad (7)$$

where $\mathrm{n_{gi}}$ represents the group indices of the five propagating modes, Eff_{gi} means the Effective group indices of the five propagating modes and λ is the wavelength of the five propagating modes The DMGD for the neighboring modes are expressed by equation (8) (Linqi *et al.,* 2022).

$$\mathrm{DMGDij} = (L \times 10^3 / c) \times (\mathrm{n_{gi}} - \mathrm{n_{gj}}) \qquad (8)$$

where DMGDij is the deferential mode group delay of the propagating ith and jth mode, L represent the length of the MCF, c means the speed of light in vacuum, $\mathrm{n_{gi}}$ denote the group indices of the ith propagating mode and $\mathrm{n_{gj}}$ expresses the group indices of the jth propagating mode.

IM-XT for Each Neighboring Modes

Inter Mode Crosstalk for the neighboring modes are expressed by Equation (9) (Snyder & Love, 1983).

$$\mathrm{IMXT}_{ij} = 10 \times \log 10\,(\mathrm{kappa_{ij}} \times \mathrm{L})^2 \qquad (9)$$

The simulation parameters for generating the dataset for the three Index profile FMF design are presented in Table 2. It contains the parameters of the speed of light, the propagation wavelength, the numbers of FMF design, the refractive index of the cladding and their relative values respectively.

Table 2: Simulation Parameters for the Few Mode Fiber

Parameter	Value
Speed of light in vacuum	$3*10^8$ m/s
Wavelength	1550 nm
Number of designs	10000
Refractive index of cladding	1.44

The flowchart for generating the 3-index profile design parameters for the FMF is presented in Figure 4. It states the procedures followed to develop the dataset for the FMF parameters.

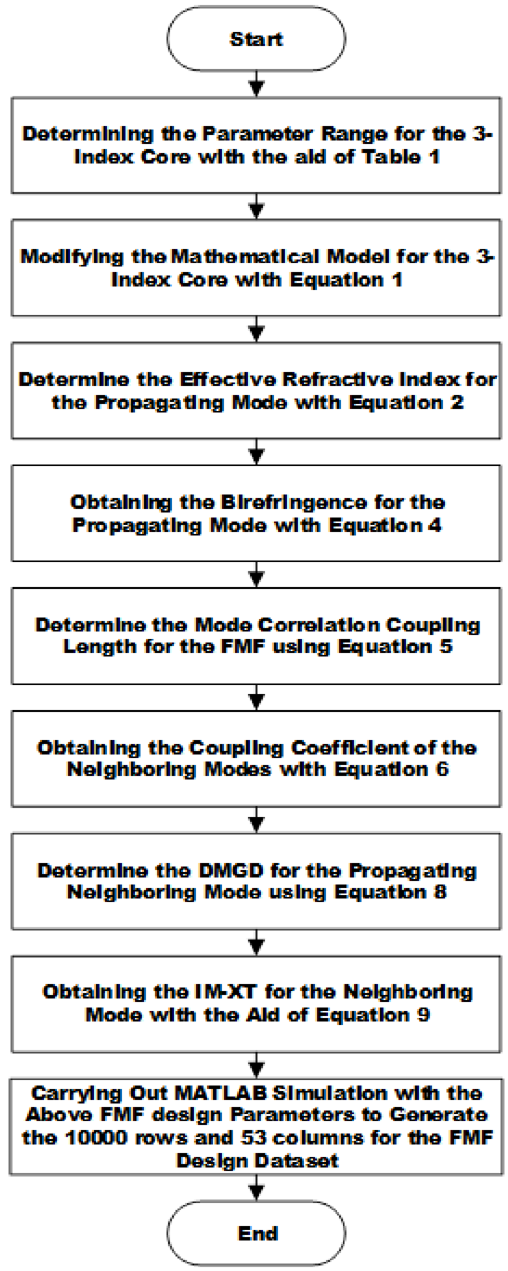

Figure 4. Generating 3-index profile design parameters for the FMF

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Result of the Three Index Profile FMF Design Dataset

The generated dataset was actual with the aid of Equations (1) to (9) with additional design parameters when compared with the dataset used in exist work. The additional design parameters, variables and features are; the birefringence of the FMF design, the mode coupling correlation length of the neighboring mode, the coupling coefficient of the neighboring modes, the DMGD of each neighboring mode in the FMF and the IM-XT of the adjacent mode in the FMF design. The dataset was generated using the above mathematical equations with the aid of MATLAB software. From the parameters expressed above the dataset was generated having 10000 rows and 53 columns with step size of 1×10^{-4} as shown in the FMF Design Dataset. csv.

Dataset Validation for the Few Modes Fiber Design Parameters

The validations of the FMF design dataset are presented in this section. It consists of the results of the correlation matrix of the FMF design dataset, the boxplot for the outlier detection of the FMF design parameters and the Principal Component Analysis (PCA) variance of the design dataset are presented respectively.

Result for the Correlation Matrix of the Few Modes Fiber Design Dataset

The correlation matrix shows how strongly design parameters of the FMF dataset are related. Figure 5 provides the Pearson Correlation Matrix (PCM) heatmap, the tool is used for data analysis and machine learning to visualize how parameters, variables and features in a dataset are linearly related to each other. PCM is useful for detecting redundant features that carry similar information. The correlation coefficient ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation). The color map with blue color intensity represents the correlation value. The dark blue represents strong correlation for either positive or negative. These indicates that, as one parameters, variables and features increase, the other tends to increase as well or one variable increases as the other decreases.

Features with strong negative correlation suggest inverse relationships between parameters, variables or features. While the white color (near 0) means weak correlation or no correlation. These variables are linearly independent; this is important for model diversity. Such features contain unique information for valuable parameters, variables and feature prediction. The x and y axes labels are for parameters, variables and features in the FMF dataset for the 3-index profile core parameters. Though labels are overlapping, they represent individual parameters, variables and features in the dataset of the FMF parameters. The diagonal line shows the correlation of the parameters, variables and features with itself are always 1, so the diagonal will always show perfect correlation (deep blue). The symmetry is the matrix which are symmetrical about the diagonal because the correlation between parameter x and parameter y is the same as between y parameter and x parameter. This allows for quantifying and validating the correlation strength the generated dataset of the few mode fiber parameters.

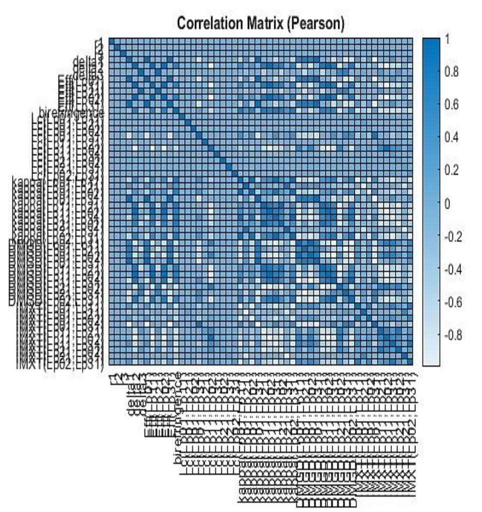

Figure 5: Correlation Matrix Heatmap for the 3-Index Profile Core of the FMF

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

The percentage correlation matrix for the 3-index profile parameters of the FMF design are shown by Equation (10) (Hall, 1999).

$$P_{[x,y]} = (\frac{C_{[x,y]}}{T}) \times 100\% \quad (10)$$

where $P_{[x,y]}$ is the percentage of the values in range [x, y], $C_{[x,y]}$ represents the number of correlation values in the range [x, y] and T coincides with the total number of entries in the correlation matrix including diagonal. The total number of entries in the correlation matrix including diagonal are expressed by Equation (11) (Hall, 1999).

$$T = N \times N \quad (11)$$

where, T coincides with the total number of entries in the correlation matrix including diagonal and N is the number of variables. Table 3 presents the results of the estimated correlation distribution for the 3-index profile of the FMF design parameters. The results were obtained with the aid of Equations 10, 11 and MATLAB software. From Table 3, it shows that, for very high positive entries parameters are 10.5%, high positive entries parameters are 12%, in the case of moderate positive correlated data are 15%, as for moderate negative correlated data is 6%, high negative are 3%, while very high negative correlations are 1% and weak/no correlation data are 52.5%.

Table 3: Results of the Estimated Correlation Distribution for the FMF

Correlation Range	Strength Description	Total Entries (%)
+0.9 to +1.0	Very High Positive	10.5
+0.7 to +0.9	High Positive	12
+0.5 to +0.7	Moderate Positive	15
-0.5 to -0.7	Moderate Negative	6
-0.7 to -0.9	High Negative	3
-0.9 to -1.0	Very High Negative	1
-0.3 to +0.3	Weak / No Correlation	52.5

Boxplot for the Outlier Detection of the FMF Dataset

The boxplot in Figure 6 presents the outliers detected in the dataset for the 3-index profile design of the FMF. The x-axis is the design parameters, variables or features of the 3-index profile of the FMF. While the y-axis is the numbers of outliers of the 3-index profile of the FMF dataset. The central box shows the Inter-Quartile Range (IQR) which range from Q1 (25th percentile) to Q3 (75th percentile). The line inside the box is the median (Q2) of the boxplot. The "whiskers" extend to the lowest and highest values within 1.5×IQR of the box. The red crosses (squares) indicate outliers' points beyond 1.5×IQR from the box. The plot shows multiple boxplots side-by-side for many parameters, variables or features. Boxplot for outlier detection implies this was generated to visualize or flag abnormal values across multiple variables. The x-axis consists of the of the 3-index profile parameters of the FMF.

Witch begins with the radian distance of the 3-index core and ends with the IMXT of the neighboring modes in the fiber core. The y-axis scale begins from zero and goes up to $2×10^4$, indicating a wide spread in some parameters, variables or features. The unit scale varies widely, suggesting strong feature scaling or data normalization. Outliers with red markers means some features have a large number of outliers both on the high end (above whiskers) and the low end (below whiskers, near 0). The outliers near 0 means zero inflation (many zeros in features) log-normal or exponential data skew possible missing values encoded as 0. Central features with many boxplots cluster around the lower middle range of the y-axis. Median lines are mostly centered inside the boxes, indicating reasonable symmetry in many features. The outlier density expresses few

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

features on the left and right edges have extremely high density of outliers.

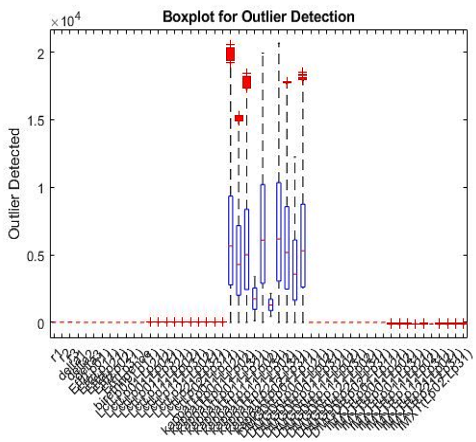

Figure 6: Boxplot Outlier for the FMF Dataset

Table 4 presents the result of the percentage outlier for the FMF dataset. Table 4 consists of features (parameters or variables) of the 3-index profile of the FMF and the percentage of detected outlier of the 3-index profile of the FMF as well as the percentage normal features (parameters or variables) of the FMF. The outliers are widespread, affecting multiple features of the FMF dataset. The clustering of the medians and the boxes in the plot is tight and the range with widely scattered outliers shows the majority of the data are in centered, but a few extreme values distort the scale. The result of the percentage outlier and normal data for the FMF dataset are presented in Table 4. The result in Table 4 was actualized with the aid of MATLAB software.

Table 4: Result of the Percentage Outlier and Normal Data in the Few Modes Fiber Dataset

Feature	Outlier %	Normal %
F1	0.0	100.0
F2	0.0	100.0
F3	0.0	100.0
F4	0.0	100.0
F5	0.0	100.0
F6	0.0	100.0
F7	0.0	100.0
F8	0.0	100.0
F9	0.0	100.0
F10	0.0	100.0
F11	0.0	100.0
F12	0.0	100.0
F13	11.7	88.3
F14	12.36	87.64
F15	11.98	88.02
F16	10.86	89.14
F17	12.13	87.87
F18	3.05	96.95
F19	12.29	87.71
F20	12.05	87.95
F21	11.99	87.95
F22	12.41	87.59
F23	0.34	99.66
F24	0.1	99.9
F25	0.3	99.7
F26	0.0	100.0
F27	0.0	100.0
F28	0.0	100.0
F29	0.0	100.0
F30	0.03	99.97
F31	0.0	100.0
F32	0.13	99.87
F33	0.0	100.0
F34	0.0	100.0
F35	0.0	100.0
F36	0.0	100.0
F37	0.0	100.0
F38	0.0	100.0
F39	0.0	100.0
F40	0.0	100.0
F41	0.0	100.0
F42	0.0	100.0
F43	4.03	95.97
F44	3.7	96.3
F45	3.89	96.11
F46	2.68	97.32
F47	4.16	95.84
F48	0.0	100
F49	4.12	95.88
F50	3.91	96.09
F51	3.46	96.54
F52	4.0	96.0

The overall average percentage outlier data in the FMF dataset is 2.80% was obtained by Equation (12) (Chandola *et al*., 2009).

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

$$\text{Average Outlier }\% = \frac{\sum \text{Outlier}\%}{\text{Total Number of Outliers}} \quad (12)$$

where, average outliers % means the summation of the percentage outlier divided by the total number of outliers. $\text{Outlier }\%$ represents the percentage outlier for each parameter in the dataset and the number of outliers denotes the total numbers of outlier in the dataset. Likewise, the overall average percentage normal data in the FMF dataset is 97.20% was determine with Equation (13) (Chandola *et al*., 2009).

$$\text{Average Normal }\% = \frac{\sum \text{Normal}\%}{\text{Total Number of Normal}} \quad (13)$$

Principal Component Analysis Variance

The Principal Component Analysis (PCA) variance for the 3-index profile dataset of the FMF design is presented in Figure 7. The PCA variance plot shows how much parameters, variance or features for each Principal Component (PC) retains from the original dataset varies from one another. The bar chart represents the variance by each PC for the percentage variance of the PCA of the 3-index profile of the FMF dataset. While that of the righthanded y-axis presents a linear graph for the percentage cumulative variance for the three PC in the FMF dataset. From the PCA bar chart, the x-axis consists of PC1, PC2 and PC3 respectively. While the y-axis at the left hand is the percentage variance and the y-axis at the right is the percentage cumulative variance for the PCA of the FMF dataset.

The simulation results for the PCA variance were implemented with the aid of MATLAB simulation software. From the result it shows that the order of descending in the magnitude of the percentage variance for the left side y-axis of the PCA plot. The percentage variance for the principal component PC1 is 63.9795%, that of PC2 is 19.3073% and that of the PC3 is 13.2516% respectively. For the right-hand side of the y-axis in the PCA plot, there was a linear increment in the ascending order for the percentage cumulative variance for the PCA component. The simulation result for the percentage cumulative variance for PC1 is 63.9795%, that of PC2 is 83.2868% and that of PC3 is 96.5384% respectively.

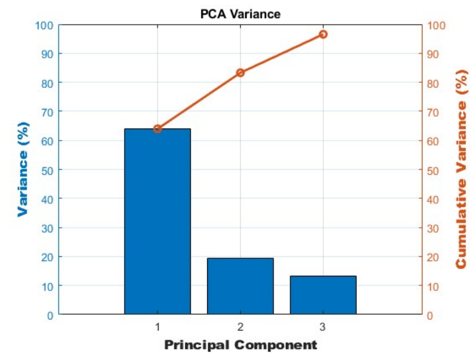

Figure 7: Principal Component Analysis Variance

CONCLUSION

This research work presents the development of three index profile core dataset for FMF design parameters for training machine learning models for long haul optical communications applications. The generated dataset for the FMF design was developed for long haul optical fiber with length starting from 100Km to 1000Km. The FMF design parameters used for developing the dataset consists of the radian distance of the 3-index profile core, the difference between the refractive indices and that of the index of the cladding and the effective refractive index of the propagating modes. Furthermore, additional design properties for the FMF were determine they are; the birefringence of the propagating modes, the coupling correlation lengths of the optical length, the coupling coefficient of the propagating modes, the DMGD and the IM-XT for the five guides modes for the FMF design.

The realized FMF design dataset was generated using mathematical equations that characterize FMF design parameters and with the aid of MATLAB software. The dataset of the FMF design consists of 10000 rows and 53 columns with step size of 1×10^{-4}. The simulation analysis carried out for the 3-index FMF design dataset validation shows that, for the correlation matrix analysis, for very high positive entries the value is 10.5%, for high positive entries the value is 12%, in the case of moderate positive correlated data

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

the value is 15%, as for moderate negative correlated data the value is 6%, likewise for high negative data the value is 3%, while for very high negative correlations the value is 1% and for weak/no correlation data the value is 52.5%. Considering the boxplot analysis for the percentage outlier and normal data where examined. The overall average percentage for outlier data is 2.80% while that of the normal data in the FMF dataset is 97.20%.

The analysis for the PCA variance were also discussed while includes the percentage variance, the PC and the percentage cumulative variance. From the result it shows that the order of descending in the magnitude of the percentage variance for the left side y-axis of the PCA plot. The percentage variance for the principal component PC1 is 63.9795%, that of PC2 is 19.3073% and that of the PC3 is 13.2516% respectively. For the right-hand side of the y-axis in the PCA plot, there was a linear increment in the ascending order for the percentage cumulative variance for the PCA component. The simulation result for the percentage cumulative variance for PC1 is 63.9795%, that of PC2 is 83.2868% and that of PC3 is 13.2516% respectively. With the aid of the developed dataset, machine learning models can be trained to predict different optimized design parameters for FMF therefor, minimizing design complexities and improving design precision in FMF design fabrication.

REFERENCES

Agrawal, G. P. (2010) *Fiber-Optic Communication Systems*, Wiley-Interscience, 4th Edition, 2010.

Arun, K. V. & John C. M. (2019). Fundamentals of Inorganic Glasses (Third Edition) pp 537-594.

Bhagyalaxmi, B., Gyana, R. P., Shailendra, K. V., & Mihir, N. M. (2023). Machine Learning-based Inverse Model for Few-Mode Fiber Designs. Computer Systems Science & Engineering. 45(1):311-328. DOI: 10.32604/csse.2023.029325.

Chandola, V., Banerjee, A., & Kumar, V. (2009). *Anomaly detection: A survey*. ACM Computing Surveys (CSUR).

Hall, M. A. (1999*). Correlation-based Feature Selection for Machine Learning.* researchcommons.waikato.ac.nz

Junling Hu, Hongwei Li, Hailiang Chen, Sa Zhang, Ruyue Shi, Yuxin Li, Meiyu Cai, and Shuguang Li (2024). Forward design method for the design of panda polarization-maintaining few-mode optical fiber based on artificial neural network. Vol. 32, No. 21 / 7 Oct 2024 / *Optics Express* 36848.

Ken-ichi K. & Nikolaos-Panteleimon, D., (2017). Few-Mode Optical Fibers: Original Motivation and Recent Progress. *IEEE Communications Magazine.* DOI: 10.1109/MCOM.2017.1600876

Linqi, W., Zhang, X., Liu, Y.*, et al*. (2022). *Mode coupling and differential mode group delay analysis in few-mode fibers for SDM systems*. Journal of Lightwave Technology, 40(12), 3456–3465.

Liu, S. & Xiang, L. (2024). Inter-Mode Crosstalk Estimation between Cores for LP_{mn} Modes in Weakly Coupled Few-Mode Multicore Fiber with Perturbations. *Sensors* 2024, *24*(18), 5969; https://doi.org/10.3390/s24185969

Luca, P. (2025). Mode Coupling in Optical Fibers. May 2025. Journal of Lightwave Technology. PP(99):1-12. DOI: 10.1109/JLT.2025.3543623

Mizuno, T. & Miyamoto, Y. (2017). High-capacity dense space division multiplexing transmission. OpticalFiberTechnology35(2017)108–117.

Pierre, S., Marianne B-A., & Denis, M. (2014). Few-Mode Fibers for Mode-Division-Multiplexed Systems. Journal of Lightwave Technology. 32(16):2824-2829

Qiuyan, Y., Hui, Yang, R. Z., Ao, Y., Wei, B., Yuanlong, A., Jie, Z., & Hongyun, X. (2018). Core, Mode, and Spectrum

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Assignment Based on Machine Learning in Space Division Multiplexing Elastic Optical Networks. *IEEE Access,* 6, 15898-15907.

Rademacher, G., Luis, R. S., Puttnam, B., Ryf, R., van der Heide, S., Eriksson, T. A., Fontaine, N. K., Chen, H., Essiambre, R. J., Awaji, Y., & Furukawa, H. (2022). A Comparative Study of Few-Mode Fiber and Coupled-Core Multi-Core Fiber Transmission. Journal of Lightwave Technology, 40(6), 1590-1596.

Shreya M. (2023). Fiber Optics in Communication Networks: Trends, Challenges, and Future Directions. International Journal of All Research Education and Scientific Methods (IJARESM), Volume 11, Issue 7

Snyder, A.W., & Love, J. D. (1983). *Optical Waveguide Theory*. 3(3), pp. 378-379

Xinyi, C., Jiangbing, D., Baining, Y., Ke, X., & Zuyuan, H. (2022). "Active learning aided four-mode fiber design with equalized zero dispersion for short-reach MDM optical communications," J. Opt. Soc. Am. B 39, 1958-1963

Zhiqin, H., Jiangbing, D., Xinyi, C., Weihong, S., Yuting, H., Chang W., Ke X., & Zuyuan, H. (2020). Machine learning aided inverse design for few-mode fiber weak-coupling optimization. *Optics Express*, 28(15).

Corresponding author: S. P. Obaje
objsunpet948@gmail.com
Electronics and Telecommunications Engineering Department, Ahmadu Bello University, Zaria, Nigeria.