data

MDPI

Data Descriptor

Multi-SpeckleForce: An Experimental Multimode Fiber Specklegram Dataset for Multi-Position Force Sensing

Rabiul Al Mahmud [1,2], Naveed Iqbal [2,3] and Khurram Karim Qureshi [1,2,3,*]

[1] Optical Communication and Sensing Laboratory (OCSL), King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia
[2] Electrical Engineering Department, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia
[3] Interdisciplinary Research Center for Communication Systems and Sensing (IRC-CSS), King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia
* Correspondence: kqureshi@kfupm.edu.sa

Abstract

This experimental dataset presents optical fiber specklegram images obtained from a step-index multi-mode fiber (MMF) due to multi-position force perturbations. The force recognition specklegrams can help in the development of data-driven models for simultaneous force estimation at multiple sensing locations along a single MMF. The dataset comprises 11,000 grayscale images with a resolution of 256 × 256 pixels, which belong to 550 records across five independent acquisition sets. Each record corresponds to a fixed triplet of applied transverse forces at three spatially separated sensing positions. The applied force values span a range from 0.26 N to 1.26 N and measured using digital force gauges with a readout resolution of 0.01 N. Key acquisition parameters include a 633 nm HeNe laser source, a 50 µm core multimode fiber of 2 m length, and CCD-camera-based specklegram imaging. Structured labels are provided in physical units of Newtons for all images. This dataset supports the development and systematic evaluation of multi-output regression models for specklegram-based force sensing and can facilitate comparative studies across experimental and data-driven approaches. Two baseline machine learning implementations with k-nearest neighbors (kNN) and Random Forest (RF) are provided to demonstrate the usability of the dataset for data-driven multi-position force estimation. Furthermore, this resource will enable research in applications such as robotic tactile sensing, structural health monitoring, and distributed force measurement systems.

Dataset: https://doi.org/10.5281/zenodo.19366118

Dataset License: CC-BY-SA

Keywords: specklegram; optical fiber sensor; force sensing; multimode fiber; multi-position sensing; dataset; machine learning; CCD imaging

Academic Editor: Tianwen Zhang

Received: 4 July 2026
Revised: 5 September 2026
Accepted: 9 September 2026
Published: 16 September 2026

1. Summary

Optical fiber sensing technologies have attracted significant attention over the past decades due to their immunity to electromagnetic interference, compact form factor, mechanical flexibility, and suitability for operation in harsh environments [1–7]. Among various optical fiber sensing modalities, multi-mode fiber (MMF) specklegram sensing has emerged as a promising approach because the specklegram formed at the fiber output originates from complex modal interference and is highly sensitive to external perturbations such as strain, tactile, bending, temperature, refractive index, and applied force [8–14]. Even small mechanical disturbances can induce measurable variations in the spatial intensity distribution of the speckle pattern. This inherent sensitivity enables high-resolution sensing without requiring wavelength interrogation or interferometric stabilization, as typically employed in traditional optical fiber sensing designs such as fiber Bragg gratings (FBGs), long-period gratings (LPGs), and photonic crystal fibers (PCFs) [12,15–18]. Historically, the nonlinear relationship between sensing parameters and specklegram patterns was interpreted using analytical and mathematical models. More recently, advances in deep learning have enabled direct learning of complex nonlinear mappings between speckle intensity distributions and physical parameters from experimental data with convolutional neural networks (CNNs), transformer-based architectures, and other machine learning models demonstrating a strong capability in specklegram-based sensing applications [19–24]. Beyond these sensing applications, related MMF research has advanced active wavefront control and learning-based computational inversion. Wavefront-shaping methods have achieved near-perfect focusing through MMFs [25]. Moreover, deep networks have reconstructed or classified input images from distal speckle intensity patterns, including for dynamically perturbed fibers and holographically labeled image transmission [26–28]. Collard et al. [28] added holographic labels to the input phase images, creating label-dependent variance in the output speckle patterns and enabling reconstruction of labeled images and RGB components without temporal synchronization. Learning-based specklegram sensing has also been extended to spatially distributed environmental monitoring. Inalegwu et al. detected water volumes as small as 0.1 mL and identified leak spots 1 cm apart [29]. These studies share the broader goal of extracting complex information encoded in MMF specklegrams.

Despite these developments, most specklegram-based sensing studies reported in the literature are conducted as experimental research without publicly releasing structured datasets. Existing works span both classification and regression tasks and address single-parameter as well as multi-parameter sensing scenarios [14,30–35]. However, openly accessible synthetic or experimental datasets for MMF specklegram sensing remain limited [35,36]. The absence of such datasets restricts reproducibility, comparative evaluation of machine learning models, and systematic benchmarking across different algorithmic approaches. Public release of well-structured datasets is therefore essential to advance specklegram-based sensing research and to promote broader reuse within the scientific community.

To address that need, this work presents an experimental dataset of optical fiber specklegrams acquired from a step-index MMF subjected to controlled, simultaneous force perturbations at three spatially separated sensing positions. Multi-point force estimation along a continuous fiber introduces additional complexity due to mechanical coupling between sensing locations and the high-dimensional, nonlinear characteristics of specklegram variations. In the presented dataset, forces are applied at predefined positions using precision translation stages, and corresponding ground-truth values were measured using digital force gauges. The dataset comprises grayscale specklegram images organized into records across five independent acquisition sets. Each record contains a short temporal sequence of frames captured under steady-state force conditions. The structured force labels in physical units of Newtons are provided in tabular form for all images. This dataset complements our previously published work on simultaneous multi-point force detection using specklegram analysis by making the underlying experimental data publicly available for reuse, independent validation, and reproducible comparison across future methods [34].

Two classical machine learning baselines based on kNN and RF are evaluated to demonstrate the usability of the dataset as a benchmark resource. These models were selected because they are lightweight, interpretable, and do not require specialized hardware, making them broadly accessible to the research community. The baselines are assessed under a record-aware data split to establish the reference performance for multi-position force estimation from engineered specklegram descriptors. The dataset is expected to support further research in multi-output regression, domain generalization across acquisition sessions, transfer learning across optical fiber sensing configurations, and data-driven sensing strategies. Potential application areas include robotic tactile sensing, structural health monitoring, and distributed force measurement systems.

Unlike [34], the main contributions of this work are as follows:

- First, a structured and labeled benchmark dataset for multi-point force sensing, derived from our previous study [34], is presented and made publicly available [https://doi.org/10.5281/zenodo.19366118]. The dataset is accompanied by comprehensive documentation, including data organization, naming conventions, label-file specifications, and acquisition metadata, thereby facilitating reproducible research, transparent evaluation, and independent reuse by the research community.
- Second, we provide baseline validation using two classical machine learning models (kNN and Random Forest) under a record-aware evaluation protocol as interpretable reference benchmarks.
- Third, we characterize the cross-session variability of the dataset through dimensionality-reduction analysis, revealing a session-dependent structure in the feature space that enables a systematic evaluation of domain generalization and cross-session transfer learning approaches.

The experimental acquisition reported here is the dataset underlying our earlier sensor-system study [34]. The previous study proposed and demonstrated a specklegram-based multi-position force-sensing system using a skip-connected autoencoder as its regression method. However, the present Data Descriptor publishes the images and labels as a reusable resource and adds dataset-specific documentation, classical reference baselines, and record-aware and cross-session evaluation. No model architecture or performance results from [34] are presented as a new contribution in this work. Table 1 shows the relationship between these two works.

Table 1. Relationship between Ref. [34] and the present Data Descriptor.

Aspect	**Ref. [34]**	**Present Data Descriptor**
Primary objective	Developed and demonstrated a simultaneous multi-position force-sensing system based on MMF specklegrams.	Published and documented the underlying experimental dataset as a structured and reusable benchmark.
Experimental data	Used the 11,000 specklegrams from 550 records and five acquisition sessions to evaluate the sensing system.	Publicly released the same underlying acquisition with organized images, labels, and metadata.
Role of data-driven modeling	Used a skip-connected autoencoder as the regression method for simultaneous three-position force estimation.	Used handcrafted descriptors with kNN and RF as reproducible reference baselines, including mixed-session and LOSO evaluations.
Distinct contribution	Sensor-system implementation, deep learning interrogation method, and force-estimation performance.	Dataset structure and documentation, acquisition and preprocessing details, exact splits and seeds, reproducibility files, and a cross-session benchmark.

By providing an experimentally grounded and openly reusable benchmark resource, this work aims to support reproducible research and accelerate the development of robust specklegram-based force sensing methodologies. The remainder of the paper is organized as follows. Section 2 describes the dataset contents, structure, and labeling scheme. Section 3 presents the experimental methodology and acquisition procedure. Section 4 provides the baseline machine learning evaluation. Section 5 presents user notes including intended applications, limitations, and prospective extensions.

2. Data Description

This section provides a detailed description of the dataset contents, structure, labeling scheme, and basic statistical characteristics. The dataset consists of force recognition specklegrams acquired from a multimode optical fiber under controlled multi-position force perturbations. The following description enables correct interpretation, visualization, and utilization of the dataset for machine learning-based sensing and applications.

2.1. Dataset Overview

The dataset consists of force recognition specklegrams acquired from an MMF under controlled multi-position force perturbations. The principal attributes are summarized in Table 2.

Table 2. Dataset attributes with details.

Attributes	**Description**
Total Acquisition Sets	5 (Sets A–E)
Total Records	550
Records per Set	73 (A), 100 (B), 100 (C), 127 (D), 150 (E)
Images per Record	20
Total Images	11,000
Image Format	.PNG
Image Type	Grayscale specklegrams
Image Resolution	256 $\times$ 256 pixels
Sampling Interval	2 s
Duration per Record	~38 s
Labels per Image	Three force values in Newtons
Force Range	0.26 N to 1.26 N
Force Gauge Readout Resolution	0.01 N

2.2. Representative Samples and Distribution of Force Values

Some of the representative force recognition specklegrams from the dataset have been depicted in Figure 1. Each image corresponds to a combination of applied forces at sensing positions Pos1, Pos2, and Pos3. These force values are shown below the images. The subtle redistributions of intensity encode the applied force information on the specklegram, which might not be visually distinguishable. These variations reflect the sensitivity of the MMF specklegram to external changes and highlight the suitability of the dataset for data-based force sensing.

The distribution of applied forces across the entire dataset has been depicted in Figure 2. This presents the coverage and statistical properties of force values. Figure 2a shows the force values at the three sensing positions (Pos1–Pos3) across all records. Each record represents one fixed force triplet acquired as a 20-frame sequence. The record_id is an ordinal identifier within an acquisition set and restarts for each set. The unique record key is (set, record_id). In Figure 2a, the 550 records are reordered by the total applied force, F1 + F2 + F3, only to visualize force-space coverage; the horizontal position in the plot is not the original set-specific record_id. Figure 2b presents box plots depicting the distribution of force magnitudes at each sensing position. The marginal ranges and shapes of the force distributions at Pos1, Pos2, and Pos3 primarily reflect the experimental protocol used to sample force combinations via two translation stages. The three force labels are correlated with one another, with Pearson coefficients ranging from 0.941 to 0.958 (Spearman 0.945–0.967) across the full dataset, and from 0.971 to 1.000 within individual acquisition sets. The strong correlations reflect the mechanically constrained acquisition protocol with two actuators acting on three force positions, while the exact correlation coefficients also depend on the stage configurations sampled during acquisition.

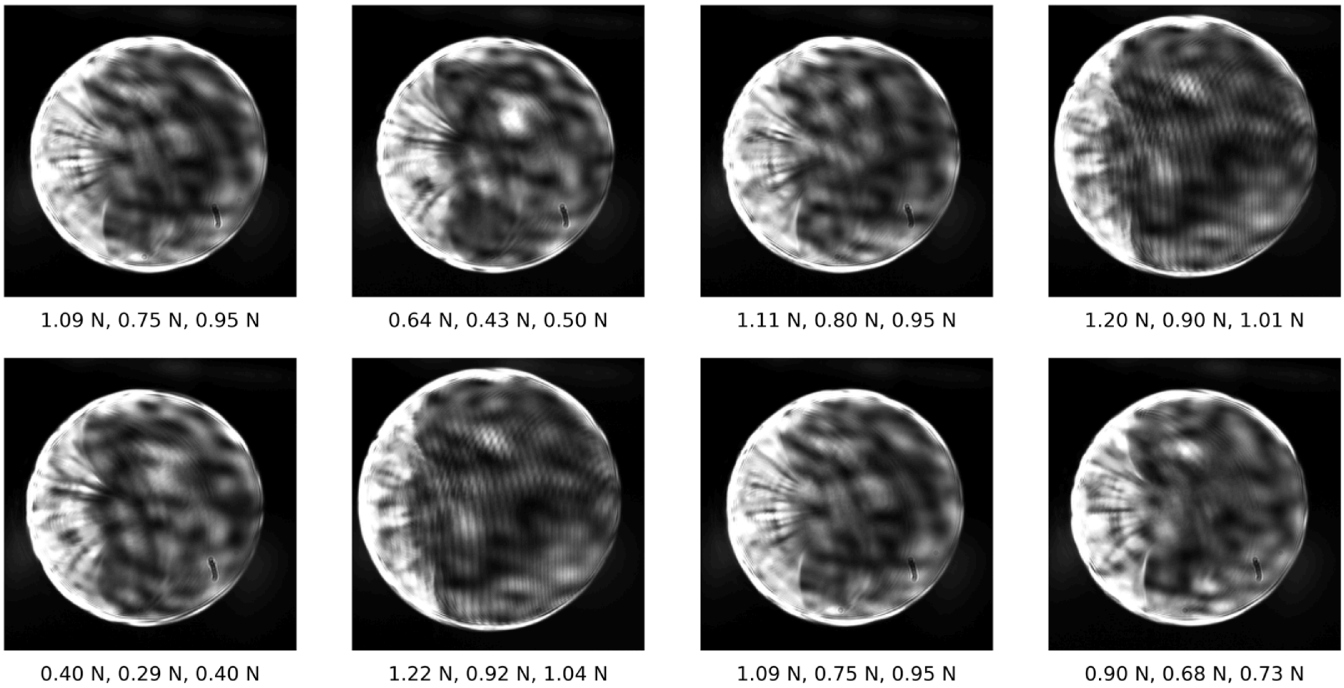
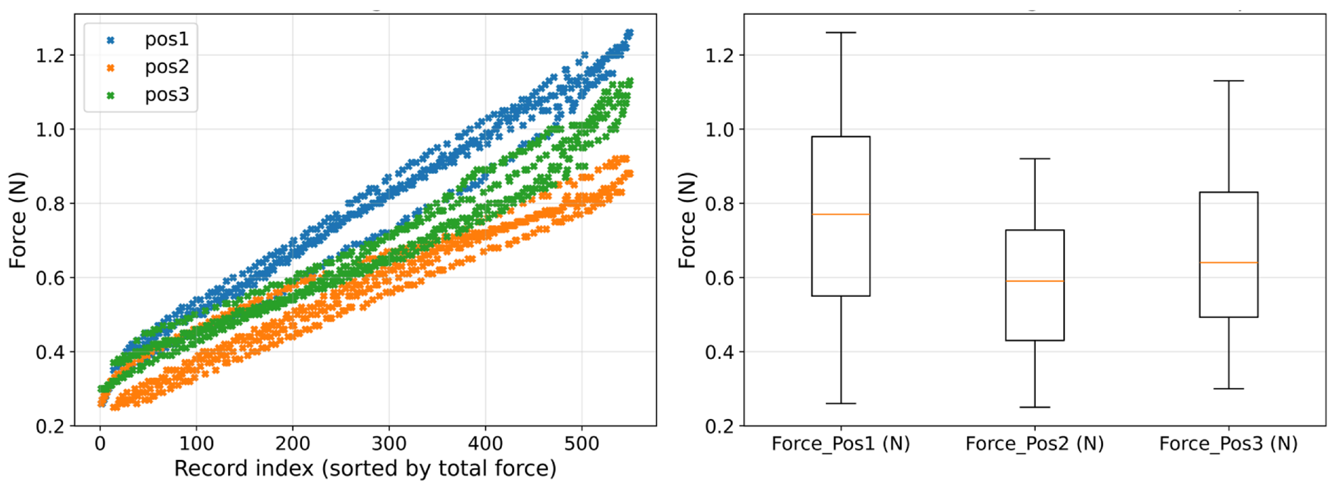


Figure 1. Force recognition specklegrams acquired under different force perturbations at sensing positions Pos1, Pos2, and Pos3.



(**a**) (**b**)

Figure 2. Dataset characterization of force labels. (**a**) Force labels for all 550 records, reordered by total applied force (F1 + F2 + F3) for visualization; each record contains one fixed force triplet and 20 frames; (**b**) Box plot summarizing the distribution of force magnitudes at each sensing position across the entire dataset.

2.3. Dataset Organization and Naming Convention

The dataset is organized in a flat directory structure, with all specklegram images stored in a single folder and corresponding labels provided in a tabular file (labels.csv). Each image filename follows a structured naming convention that encodes the acquisition set, applied forces, record index, and frame index. The general filename format with an example has been illustrated in Figure 3.

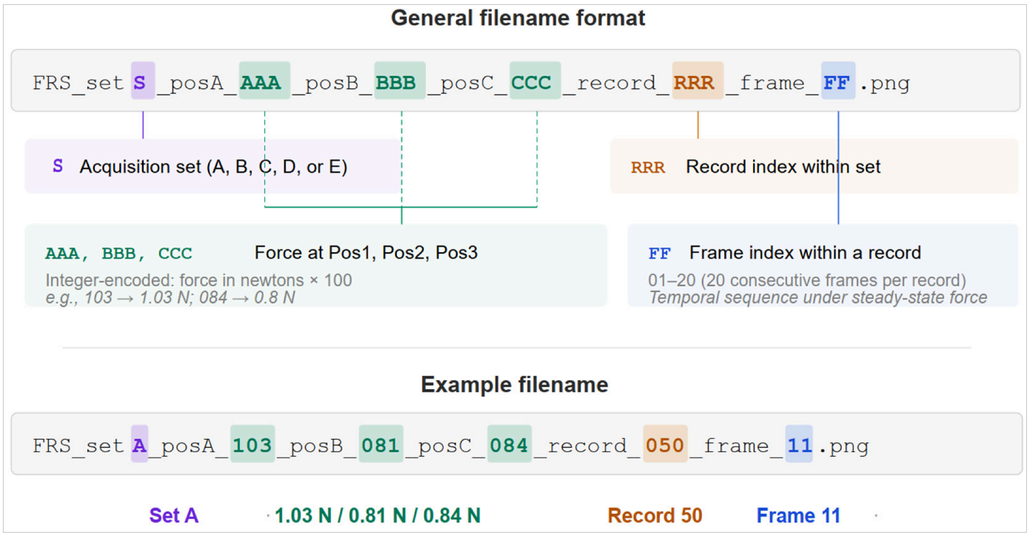


Figure 3. Dataset image file naming convention with an example.

2.4. Label File Description

The labels.csv file provides a structured mapping between the specklegram and its corresponding force values. The file contains the following columns:

- filename: Name of the specklegram image file;
- set: Acquisition set identifier (A–E);
- record_id: Record index corresponding to a fixed force condition;
- frame_idx: Frame number within a record;
- time_s: Time stamp in seconds relative to the start of the record;
- force_p1_N: Applied force at sensing position Pos1 (in Newtons);
- force_p2_N: Applied force at sensing position Pos2 (in Newtons);
- force_p3_N: Applied force at sensing position Pos3 (in Newtons).
- Digital Force gauge readout resolution: 0.01 N.

All frames within a given record share identical force values, as the applied forces remain constant during the acquisition of each 20-frame sequence. This structure enables both frame-based and record-based analysis of the dataset.

3. Methods

This section describes the experimental configuration and acquisition protocol used to generate the specklegram dataset. The objective is to present the hardware components, mechanical geometry, force application technique and acquisition procedure to enable reproducibility. The experimental setup is depicted schematically in Figure 4, showing the implemented sensing system mounted on an optical table.

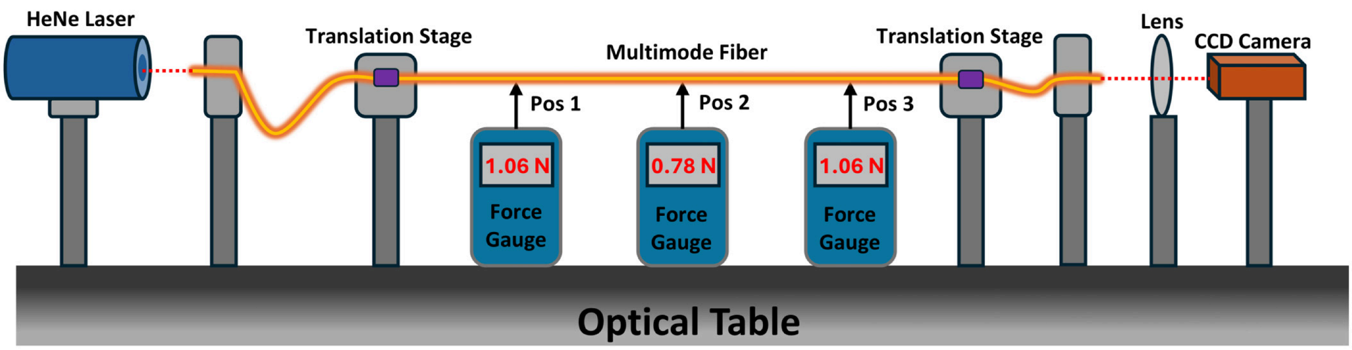


Figure 4. Schematic of the experimental configuration for multi-position force-induced specklegram acquisition, showing the 633 nm He-Ne laser source, free-space coupling into the MMF, translation stages defining the sensing span, and three force application points with digital force gauges, projection lens, and CCD camera. This experimental configuration is adapted from our previous setup in [34].

3.1. Experimental Protocol Overview

The dataset was generated by recording optical specklegrams from an MMF due to the application of forces at three fixed spatial locations. Each data sample belongs to a record. A record is defined as a temporally coherent sequence of specklegram images acquired under a fixed triplet of force values.

The acquisition protocol followed these steps:

1. A coherent laser source was launched into the multimode fiber using a free-space optical coupling arrangement. Alignment was optimized at the beginning of each acquisition session and kept unchanged during the recording of each record.
2. The fiber was mounted horizontally between two precision translation stages (TSs) to define a mechanically stable sensing span.
3. Transverse forces were applied at three sensing positions (Pos1–Pos3) by adjusting the two translation stages. For each force configuration, the forces were adjusted within the target range and held constant for a record.
4. The applied forces were simultaneously measured using digital force gauges. The measured values were assigned as ground-truth labels for all images within the corresponding record.
5. A sequence of 20 consecutive specklegram images was captured at a 2 s sampling interval, resulting in an approximate 38 s duration per record, keeping the force configuration static.
6. The procedure was repeated across multiple force combinations and across five independent acquisition sessions (Sets A–E). Between sessions, the fiber was manually repositioned, and the optical coupling alignment was re-established.

This protocol ensures that each record represents a steady-state mechanical condition, with its associated force values correspond directly to the measured applied forces.

3.2. Optical Source and Fiber Subsystem

A continuous-wave laser was employed as a light source coupling with the MMF, and the resulting modal interference at the fiber output generated a specklegram sensitive to applied force perturbations. The optical source and fiber parameters used during data acquisition are summarized in Table 3.

Table 3. Optical source and multimode fiber parameters.

Component	Specification
Laser type	Helium–Neon (HeNe)
Wavelength	633 nm
Output power	10 mW
Coupling method	Free-space optical coupling
Fiber type	Step-index MMF
Core diameter	50 μm
Cladding diameter	125 μm
Numerical aperture	0.22
Fiber length	2 m
Fiber coating	Standard protective jacket

The laser-to-fiber free space coupling alignment was optimized at the beginning of each acquisition session and kept fixed during the recording of individual data records.

3.3. Mechanical Mounting and Force Application Geometry

The multimode fiber was horizontally suspended between two precision translation stages, which defined the active sensing span. The mechanical layout and spatial geometry of the force application points are summarized in Table 4.

Table 4. Mechanical configuration and force application geometry.

Parameter	Value
Distance between translation stages	700 mm
Number of sensing positions	3
Position Pos1 from left TS	150 mm
Position Pos2 from left TS	350 mm
Position Pos3 from left TS	550 mm
Force application direction	Transverse to fiber axis

Transverse forces at Pos1, Pos2, and Pos3 were generated by adjusting the same two translation stages (say S1 and S2) that define the 700 mm sensing span. Displacing S1 and/or S2 deflects the suspended fiber and changes its tension along the entire span. The resulting local transverse force at each of the three fixed positions was measured independently using the digital force gauge mounted at that position. Because the span between S1 and S2 is a single continuous elastic structure with only two external actuation degrees of freedom, the three measured force values are not free to vary independently. Adjusting S1 and S2 changes the deflection profile of the entire span, so the forces recorded at Pos1, Pos2, and Pos3 co-vary.

Two additional fiber holders secured the MMF sections outside the 700 mm sensing span to suppress unintended motion. Within each acquisition session, these outside-span sections were maintained in a nominally fixed configuration. Their static curvature may affect the specklegram baseline, whereas any change introduced during fiber repositioning between sessions may contribute to session-to-session variability. Because the outside-span curvature was not varied independently, its isolated effect cannot be quantified from the present dataset. A close-up of the force-sensing region showing the fiber mounting and force application arrangement is presented in Figure 5.

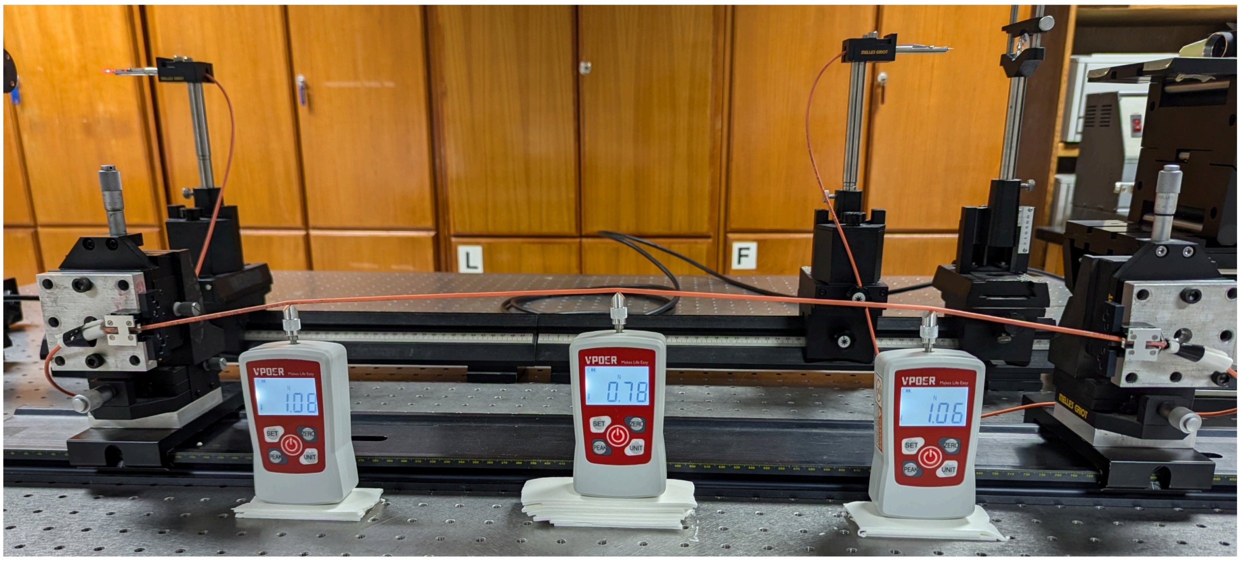


Figure 5. Close-up image of the force-sensing region showing the multimode fiber horizontally suspended between translation stages, with transverse force applied at three sensing positions using digital force gauges. This experimental configuration is adapted from our previous setup in [34].

3.4. Force Measurement System

Forces applied at the three sensing positions were directly measured using digital force gauges. The force measurement hardware and operating range are summarized in Table 5.

Table 5. Force measurement hardware and parameters.

Parameter	**Specification**
Force gauge model	VPOER Digital Force Gauge (VDFG-50) [Yueqing, China]
Readout resolution	0.01 N
Manufacturer-specified accuracy	±1% of reading
Measured force range	0.26 N–1.26 N
Number of force channels	3 (Pos1, Pos2, Pos3)

During the acquisition of each record, the applied force configuration was held constant. Mechanical coupling between sensing positions may occur due to the continuous nature of the fiber. However, the recorded labels correspond to the independently measured force values at each force gauge.

3.5. Specklegram Imaging System

Specklegrams generated at the output facet of the multimode fiber were imaged onto a CCD camera using a fixed optical projection arrangement. The imaging hardware parameters are listed in Table 6.

Table 6. Specklegram imaging and camera parameters.

Parameter	**Specification**
Camera type	CCD camera
Camera model	FLIR Forge 5GigE (FG-P5G-50S4M) [Richmond, BC, Canada]
Native resolution	2448 × 2048 pixels
Acquisition mode	Continuous
Imaging optics	Fixed projection lens

The native 2448 × 2048 grayscale frames were converted to the released 256 × 256 images using a two-stage procedure designed to retain and center the circular specklegram region of interest. First, each frame was resized to 306 × 256 pixels using Lanczos interpolation, preserving the native aspect ratio exactly. The image was then cropped using a box removing 15 pixels from the left and 35 pixels from the right without vertical cropping. The resulting images were saved as 8-bit grayscale PNG files. No background subtraction, flat-field correction, or intensity normalization was applied to the released images. The baseline analysis reads pixel values directly in the 0–255 range. The final resized 256 × 256 pixels provide a standard square input for machine learning and deep learning workflows while retaining the centered circular specklegram ROI.

3.6. Multi-Session Data Acquisition Procedure

Data collection was performed across five independent acquisition sessions, labeled Sets A–E. Each session consisted of multiple force configurations recorded using identical hardware settings. Between successive acquisition sessions, the fiber was manually repositioned, and re-aligned. The laser-to-fiber coupling alignment was re-established before the start of each new session. A time interval of approximately 90 min separated Sets A–D, and Set E was acquired on a different day. No post-alignment correction or compensation was applied during data recording.

This acquisition strategy introduces session-to-session variability due to the mechanical repositioning and coupling conditions, reflecting practical deployment scenarios for multimode fiber sensing systems.

3.7. Environmental Conditions and Monitoring

All experiments were conducted in a laboratory environment under controlled ambient conditions. The environmental parameters monitored during data acquisition are summarized in Table 7.

Table 7. Environmental conditions during data acquisition.

Parameter	Condition
Ambient temperature	24 ± 1 °C
Temperature monitoring	Local thermometer near setup
Ambient light	Minimized
Fiber support	Horizontally suspended
External vibration control	Stable mounting and clamping

No active environmental compensation was applied during acquisition. The dataset therefore captures specklegram variations under controlled but realistic laboratory conditions. Temperature was monitored but was not varied as an independent experimental parameter. MMF speckle patterns can change with temperature because thermo-optic and thermo-mechanical effects modify the modal phase distribution. Consequently, the present dataset should be interpreted as a force-sensing dataset acquired within 24 ± 1 °C. It does not permit the independent temperature sensitivity of the specklegrams to be quantified.

4. Baseline Evaluation of the Dataset

This section presents a baseline machine learning evaluation to assess the usability of the specklegram dataset for data-driven multi-position force estimation. The evaluation is based on handcrafted descriptors and two classical regression models—kNN and RF. These models were selected because they are lightweight, interpretable, and do not require GPU resources, making them broadly accessible for benchmarking purposes. Each measurement record contains 20 consecutive specklegram frames acquired under the same force triplet at

Pos1, Pos2, and Pos3. Both frame-level descriptors and record-level aggregations are used to evaluate prediction performance at two granularities.

4.1. Feature Extraction

At the frame level, 19 descriptors were extracted from each specklegram image and organized into two families. The first family consists of 16 statistical and texture descriptors: mean, standard deviation, minimum, maximum, percentiles (p10, p25, p50, p75, p90), interquartile range, entropy, energy, and four gray-level co-occurrence matrix (GLCM)-based texture measures (contrast, homogeneity, energy, and correlation). The second family consists of three speckle-domain descriptors: speckle contrast, power-spectrum centroid, and grain half-maximum radius.

At the record level, the 20-frame sequence was summarized into 44 descriptors grouped into three families. The first family comprises 32 aggregated statistical and texture descriptors, obtained from the mean and standard deviation of the 16 frame-level statistical features across the 20 frames in each record. The second family comprises six aggregated speckle-domain descriptors, obtained similarly from the mean and standard deviation of the three frame-level speckle features. The third family consists of six structural similarity index (SSIM)-based temporal descriptors that capture frame-to-frame stability within each record. This two-level representation enables evaluation at both individual-image and sequence-aggregated granularities, supporting both frame-wise and record-wise modeling approaches.

For a compact formulation, let $\mathbf{x}_{r,t} \in \mathbb{R}^{19}$ denote the frame-level descriptor vector for record r and frame t, where $t = 1, \ldots, 20$. The corresponding record-level descriptor is

$$\mathbf{z}_r = [\boldsymbol{\mu}_t\ (\mathbf{x}_{r,t}), \boldsymbol{\sigma}_t\ (\mathbf{x}_{r,t}), \mathbf{q}_r] \in \mathbb{R}^{44} \tag{1}$$

where $\boldsymbol{\mu}_t(\cdot)$ and $\boldsymbol{\sigma}_t(\cdot)$ are applied feature-wise over the 20 frames and $\mathbf{q}_r \in \mathbb{R}^6$ contains the SSIM mean, standard deviation, minimum, final value, temporal slope, and normalized area under the SSIM-time curve.

4.2. Classical Machine Learning Baselines

Two classical regression models were selected as baseline predictors for multi-position force estimation: kNN and RF. The two models represent complementary learning strategies. The kNN regressor is a locality-based model in which a test sample is predicted from the target values of the most similar training samples in the feature space. In the present evaluation, a StandardScaler was fitted only on the training descriptors within a scikit-learn Pipeline, followed by distance-weighted kNN with k = 7. RF constructs an ensemble of decision trees trained on bootstrapped subsets of the data and random subsets of the input features and combines their outputs through averaging. This makes the model less dependent on a single partition of the feature space and better able to capture nonlinear interactions among heterogeneous descriptors. The RF model used the unstandardized descriptors and 400 trees with square-root feature sampling at each split.

Let $\mathrm{f_r} = [\mathrm{F_{1,r}}, \mathrm{F_{2,r}}, \mathrm{F_{3,r}}]_\mathrm{T}$ be the three-output force label. For kNN, distance is evaluated in the standardized feature space, and $\mathrm{N_7(z_r)}$ denotes the seven nearest training records. For RF, $h_\beta(\cdot)$ is the β-th regression tree. The two baseline predictions are

$$\hat{f}_r(kNN) = \left[\sum\nolimits_i N_7(z_r) d_{ri}{}^{-1} f_i\right] / \left[\sum\nolimits_i N_7(z_r) d_{ri}{}^{-1}\right] \tag{2}$$

$$\hat{f}_r(RF) = \frac{1}{400} \sum\nolimits_{\beta=1}^{400} h_\beta\ (\mathbf{z}_r). \tag{3}$$

Here, d_{ri} is the Euclidean distance between standardized feature vectors. At the frame level, the same mapping is used with $\mathbf{x}_{r,t}$ in place of $\mathbf{z}_r$.

The benchmark evaluation was carried out at both the frame level and record level. At the frame level, each individual specklegram is treated as one sample, yielding 11,000 frame instances in total. These were split into 7680 training samples, 1660 validation samples, and 1660 test samples, corresponding to approximately 70%, 15%, and 15%, respectively. At the record level, the 20-frame sequences were summarized into one aggregated descriptor vector per measurement record, resulting in 550 total records. These were split into 384 training records, 83 validation records, and 83 test records using the same proportions. In both cases, a record-aware split was enforced so that all frames belonging to the same record remained within the same partition, preventing data leakage between training and test sets. To support computational reproduction of the baseline evaluation, the public GitHub repository https://github.com/rabiulalmahmud5/Multi-SpeckleForce (accessed on 8 September 2026) contains the feature-extraction and baseline-reproduction script, the exact (set, record_id) assignments for the training, validation, and test partitions (train_records.csv, val_records.csv, test_records.csv), the LOSO definitions, the random seed (42), software_versions.txt, requirements_exact.txt, the model configuration, the reproduced results, and the captured software environment (Python 3.9.16; NumPy 1.24.3; pandas 1.5.3; SciPy 1.10.1; scikit-learn 1.2.2; scikit-image 0.24.0; OpenCV 4.8.1.78; Pillow 11.1.0). The native-image preprocessing source code is not included. The complete deterministic procedure and numerical parameters are reported in Section 3.5.

Prediction performance was evaluated using mean absolute error (MAE) and the coefficient of determination (R^2). At the frame level, kNN achieved an MAE of 0.0117 and R^2 of 0.9951, while RF obtained an MAE of 0.0172 and R^2 of 0.9859. The frame-level MAE of 0.0117 N is numerically close to the 0.01 N gauge display increment. This value quantifies agreement between the predictions and the recorded force labels under a record-aware split of the five-session dataset. It represents within-distribution interpolation and should not be interpreted as the absolute measurement accuracy or resolution of the sensing system. The force label fidelity remains bounded by the measurement hardware, including the manufacturer-specified accuracy. This frame-level result should also be interpreted with respect to evaluation granularity. The record-aware split prevents any individual record from being divided across partitions, but the five acquisition sessions are represented in all three partitions. Because the 20 frames within a record are visually near-identical, and records from the same session share a common optical alignment established at the start of that session, a held-out test frame can closely resemble training frames drawn from the same session even though its specific record was never seen during training. This session-level similarity, rather than direct frame leakage, is a primary contributor to the low frame-level error.

At the record level, RF clearly outperformed kNN with an MAE of 0.0509 and R^2 of 0.9033, compared to an MAE of 0.1272 and R^2 of 0.5027 for kNN. Within each model, the per-position performance is of comparable magnitude, as shown in Table 8. Prediction error and explained variance are not concentrated on a single sensing position. The frame-level predicted-versus-actual responses for both models on the test set are depicted in Figure 6. In both cases, the predicted values follow the ground truth closely, confirming that the engineered descriptors preserve force-related information and that the dataset supports stable data-driven multi-position force estimation using baseline machine learning methods.

Table 8. Per-position record level performance under the mixed-session test split.

Position	kNN MAE	kNN R^2	RF MAE	RF R^2
Pos1	0.1527	0.4641	0.0625	0.8869
Pos2	0.1092	0.4884	0.0449	0.9030
Pos3	0.1196	0.5555	0.0453	0.9201

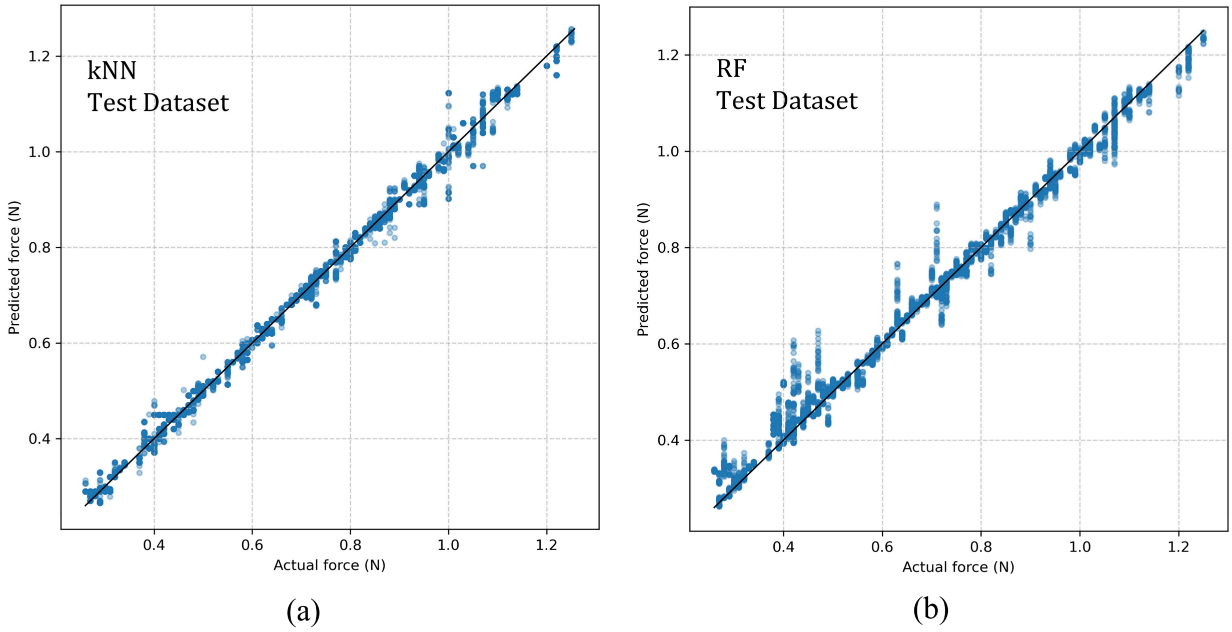


Figure 6. Frame level predicted-versus-actual force values obtained using the (**a**) kNN and (**b**) RF on the test set under the record-aware mixed-session split, showing the predicted forces at the three sensing positions (Pos1–Pos3) as a function of the corresponding ground-truth values.

4.3. Cross-Session Variability

The dataset contains five independent acquisition sessions (Sets A-E) between which the fiber was manually repositioned, and the optical coupling was re-established. This introduces session-to-session variability into the specklegram feature distributions. Cross-session generalization was evaluated using five record-level leave-one-set-out (LOSO) folds shown in Table 9, which reports the per-fold MAE and R^2. Across the five folds, mean MAEs were 0.2249 N for kNN and 0.1803 N for RF, compared with 0.1272 N and 0.0509 N under the mixed-session split. The corresponding mean R^2 was −0.752 for kNN and −0.103 for RF. Negative R^2 values indicate that the prediction errors exceeded those obtained by using the mean target value of that session as a reference for the corresponding held-out sessions. This performance gap relative to the mixed-session split quantifies substantial session-domain shift and indicates that recalibration, domain adaptation, or more session-diverse training data are needed for reliable transfer to a previously unseen acquisition session.

Table 9. Record-level leave-one-set-out cross-session performance.

Held-Out Set	kNN MAE	kNN R^2	RF MAE	RF R^2
A	0.2735	−0.961	0.2416	−0.650
B	0.1997	−1.139	0.1128	0.408
C	0.2268	−0.786	0.1946	−0.391
D	0.1680	0.034	0.1457	0.315
E	0.2563	−0.908	0.2068	−0.199
Mean	0.2249	−0.752	0.1803	−0.103

To visualize this structure, Uniform Manifold Approximation and Projection (UMAP) was applied to the record-level feature vectors. The resulting two-dimensional projection, colored by acquisition set, is shown in Figure 7. The five acquisition sets do not fully overlap in the embedding space. Different sets occupy partially separated regions, indicating that the record-level feature distribution carries a session-dependent signature. This cross-session variability is an inherent characteristic of the dataset that reflects realistic deployment conditions for multimode fiber sensing systems. Researchers using this dataset may exploit this structure to investigate domain generalization, cross-session transfer learning, or session-adaptation strategies—for example, through leave-one-set-out evaluation protocols.

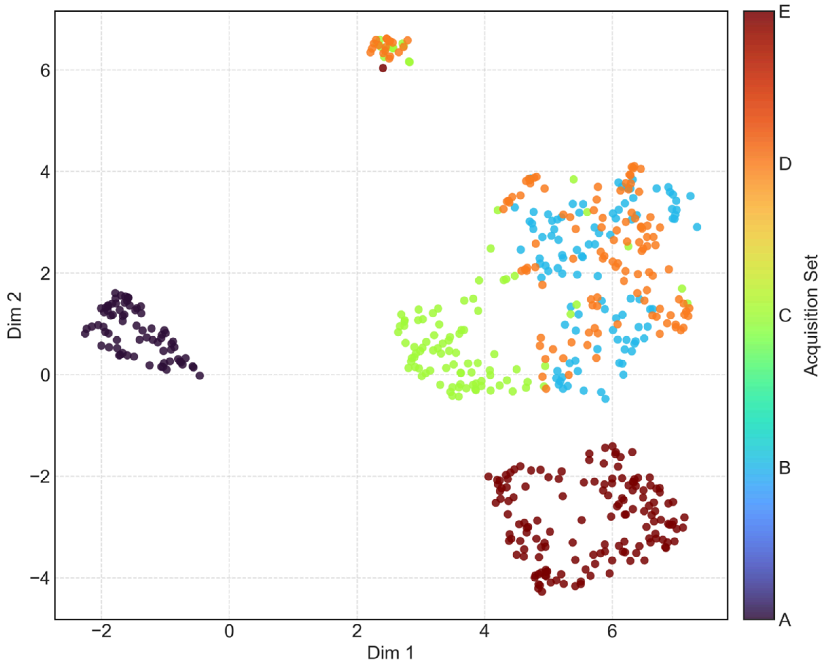


Figure 7. UMAP projection of record-level feature vectors colored by acquisition set.

5. User Notes

This section provides practical guidance for researchers intending to use the dataset, including intended applications, known limitations, and prospective extensions.

5.1. Intended Applications

The dataset is intended to support the development and benchmarking of data-driven models for simultaneous multi-position force estimation from MMF specklegrams. Each image is associated with three continuous force labels, and each record contains a temporal sequence of 20 frames acquired under a fixed force triplet. This structure enables both frame-based and record-based learning and is suitable for multi-output regression, sequence-aware modeling, feature-engineering studies, and comparative evaluation of classical machine learning as well as deep learning approaches. The five independent acquisition sessions further enable research on domain generalization, robustness to session-to-session variability, and transfer learning across related optical fiber sensing configurations. Evaluation protocols such as leave-one-set-out testing can be directly implemented using the session labels provided in the dataset.

Beyond force estimation, the dataset offers potential as a source of pre-trained deep learning representations for specklegram-based sensing more broadly. Convolutional neural networks trained on specklegram images learn hierarchical features that encode general properties of speckle patterns, including local intensity structure, texture organization, and interference-related spatial information. These learned representations are not inherently specific to force as the sensing parameter; rather, they capture how the speckle pattern responds to physical perturbations of the fiber. In a manner analogous to how models pre-trained on ImageNet provide transferable feature extractors across diverse visual recognition tasks. A model pre-trained on this specklegram dataset may serve as an initialization for sensing tasks involving other physical parameters such as temperature, bending, strain, or refractive index. Researchers working on related specklegram sensing problems could fine-tune such a model rather than training from scratch, potentially reducing data requirements and improving convergence in data-limited scenarios.

From an application perspective, the dataset is relevant to sensing scenarios in which distributed or multi-point force information is required along a compact and flexible waveguide. Potential use cases include robotic tactile sensing, where force information at multiple contact points can support manipulation and interaction tasks; structural health and load monitoring, where distributed force or deformation information is valuable for early anomaly detection; and wearable or soft-sensing systems, where flexible sensing elements are preferred over rigid sensor arrays. The dataset may also serve as a reference benchmark for comparing feature-engineered, hybrid, and end-to-end vision models for specklegram-based force sensing.

5.2. Limitations

The measurements were obtained under a fixed experimental configuration, including a 633 nm HeNe laser source, a 50/125 µm step-index multimode fiber of 2 m length, and controlled laboratory conditions. Therefore, direct transferability to other fiber types, source wavelengths, fiber lengths, or uncontrolled environmental conditions may require domain adaptation or model retraining. In addition, the three force channels are measured at distinct points along a single continuous fiber using two translation stages, so mechanical coupling between sensing positions introduced interdependence among the force responses. Although this reflects realistic physical behavior in continuous-fiber sensing, it should be taken into account when designing models that assume fully independent output channels. Furthermore, as demonstrated in Section 4.3, the five acquisition sessions exhibit meaningful variability in the feature space due to fiber repositioning and coupling re-establishment. While this variability is a valuable property for generalization studies, it also means that models trained on a subset of sessions may not generalize directly to unseen sessions without appropriate adaptation strategies.

5.3. Prospective Extensions

Future extensions of this work will focus on increasing the scope and realism of the dataset. One natural direction is to expand the sensing configuration beyond three positions so that higher-density distributed force sensing can be studied. Another is to introduce additional physical variables, such as temperature, bending, or humidity, so that the dataset can support multi-parameter sensing problems. Building a family of specklegram datasets across different physical parameters and fiber configurations would further enable the development of general-purpose pre-trained specklegram models, analogous to foundation models in computer vision, where a shared learned representation supports fine-tuning across diverse downstream sensing tasks. Increasing the diversity of acquisition conditions, including larger session variability and dynamic force variations, would also improve the value of the dataset for real-world robustness studies. These extensions would broaden the applicability of MMF specklegram sensing and support the development of more general, scalable, and deployment-ready data-driven sensing models.

Author Contributions: Conceptualization, R.A.M. and N.I.; methodology, R.A.M.; software, R.A.M.; validation, R.A.M., N.I. and K.K.Q.; investigation, R.A.M.; resources, N.I. and K.K.Q.; data curation, R.A.M. and K.K.Q.; writing—original draft preparation, R.A.M.; writing—review and editing, N.I. and K.K.Q.; visualization, R.A.M.; supervision, N.I. and K.K.Q.; project administration, K.K.Q.; funding acquisition, K.K.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the support from the Deanship of Research Oversight and Coordination (DROC) at King Fahd University of Petroleum & Minerals (KFUPM) under the Interdisciplinary Research Center for Communication Systems and Sensing (IRC-CSS) project No. INCS2611.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset presented in this study is publicly available at https://doi.org/10.5281/zenodo.19366118 (accessed on 14 June 2026) under the CC-BY-SA license. The feature-extraction and baseline-reproduction code, exact train/validation/test split definitions and indices, random seed, model configuration, reproduced results, and captured software environment are publicly available at https://github.com/rabiulalmahmud5/Multi-SpeckleForce (accessed on 8 September 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MMF	Multimode Fiber
CCD	Charge-Coupled Device
HeNe	Helium–Neon
kNN	k-Nearest Neighbors
LOSO	Leave-One-Set-Out
RF	Random Forest
ROI	Region of Interest
UMAP	Uniform Manifold Approximation and Projection
R^2	Coefficient of Determination
MAE	Mean Absolute Error

References

1. Kumar, V.; Raghuwanshi, S.K.; Kumar, S. Advances in Nanocomposite Thin-Film-Based Optical Fiber Sensors for Environmental Health Monitoring—A Review. *IEEE Sens. J.* **2022**, *22*, 14696–14707. [CrossRef]
2. Yan, J.; Wang, X.; Li, T.; Zheng, Y.; Xu, M.; Hu, Y.; Zhang, H. Hand Gesture Recognition Based on Optical Fiber Specklegram Sensor. *Opt. Express* **2025**, *33*, 10378–10387. [CrossRef] [PubMed]
3. Xie, Y.H.; Wu, C.; Li, J.; Guan, B.O. Thinned Multicore Fiber Supermode Bragg Grating and Interferometer for Temperature-Compensated Refractive Index Sensing. *J. Light. Technol.* **2023**, *41*, 4473–4480. [CrossRef]
4. Min, R.; Liu, Z.; Pereira, L.; Yang, C.; Sui, Q.; Marques, C. Optical Fiber Sensing for Marine Environment and Marine Structural Health Monitoring: A Review. *Opt. Laser Technol.* **2021**, *140*, 107082. [CrossRef]
5. Pang, Y.N.; Liu, B.; Liu, J.; Wan, S.P.; Wu, T.; Yuan, J.; Xin, X.; He, X.D.; Wu, Q. Singlemode-Multimode-Singlemode Optical Fiber Sensor for Accurate Blood Pressure Monitoring. *J. Light. Technol.* **2022**, *40*, 4443–4450. [CrossRef]
6. Leal-Junior, A.G.; Diaz, C.A.R.; Avellar, L.M.; Pontes, M.J.; Marques, C.; Frizera, A. Polymer Optical Fiber Sensors in Healthcare Applications: A Comprehensive Review. *Sensors* **2019**, *19*, 3156. [CrossRef] [PubMed]
7. Chen, G.; Li, J.; Zhu, H.; Wang, Y.; Ji, H.; Meng, F. Optical Fiber Gas Sensor with Multi-Parameter Sensing and Environmental Anti-Interference Performance. *J. Ind. Inf. Integr.* **2024**, *38*, 100565. [CrossRef]
8. Shen, B.; Li, J.; Ji, Z. Multimode Fiber Specklegram Sensor for Multi-Position Loads Recognition Using Traversal Occlusion. *Sensors* **2025**, *25*, 1737. [CrossRef] [PubMed]

9. Leal-Junior, A.G.; Frizera, A.; Marques, C.; Pontes, M.J. Optical Fiber Specklegram Sensors for Mechanical Measurements: A Review. *IEEE Sens. J.* **2020**, *20*, 569–576. [CrossRef]
10. He, L.; Chen, X.; Cheng, J.; Liu, X.; Yao, P.; Xu, L. Specklegram Evolution under Fiber-End Movement: Theory, Experiment, and High-Prediction-Accuracy Displacement Sensing. *Infrared Phys. Technol.* **2026**, *152*, 106224. [CrossRef]
11. Lyu, C.; Li, P.; Zhang, J.; Du, Y. Fiber Optic Sensors in Tactile Sensing: A Review. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 7001816. [CrossRef]
12. Newaz, A.; Faruque, M.O.; Al Mahmud, R.; Sagor, R.H.; Khan, M.Z.M. Machine Learning Enabled Multimode Fiber Specklegram Sensors: A Review. *IEEE Sens. J.* **2023**, *23*, 20937–20950. [CrossRef]
13. Vangety, N.; Majee, A.; Roy, S. Analyzing Fiber Specklegrams for Speed and Weight Recognition of Toy Model Vehicle Using Deep Learning Method. *Opt. Fiber Technol.* **2025**, *90*, 104109. [CrossRef]
14. Lu, S.; Tan, Z.; Zhang, D. Dual D-Shaped Plastic Optical Fiber for Simultaneous Measurement of Refractive Index and Temperature Based on Specklegram. *Opt. Express* **2024**, *32*, 15166. [CrossRef] [PubMed]
15. Wang, H.; Fan, S.; Meng, L.; Xia, Q.; Yuan, T.; Zhang, X.; Yuan, L. Multi-Parameter Sensor Based on Cascaded Multicore FBGs and an FPI for Bending, Temperature and Pressure Measurements. *Opt. Laser Technol.* **2024**, *175*, 110782. [CrossRef]
16. Li, X.; Wang, Z.; Hu, L.; Xia, J.; Li, S.; Kuai, Y.; Fu, W.; Cao, Z.; Wang, M.; Yu, B. Rapid and Sensitive Detection of Liver Cancer Markers Based on a Microtapered Long-Period Grating Sensor Coated with Gold Nanorods. *IEEE Sens. J.* **2025**, *25*, 26608–26615. [CrossRef]
17. Riza, M.A.; Go, Y.I.; Harun, S.W.; Maier, R.R.J. FBG Sensors for Environmental and Biochemical Applications—A Review. *IEEE Sens. J.* **2020**, *20*, 7614–7627. [CrossRef]
18. Chaudhary, V.S.; Kumar, D.; Pandey, B.P.; Kumar, S. Advances in Photonic Crystal Fiber-Based Sensor for Detection of Physical and Biochemical Parameters—A Review. *IEEE Sens. J.* **2022**, *23*, 1012–1023. [CrossRef]
19. Bender, D.; Cakir, U.; Yuce, E. Deep Learning-Based Fiber Bending Recognition for Sensor Applications. *IEEE Sens. J.* **2023**, *23*, 6956–6962. [CrossRef]
20. Ding, Z.; Zhang, Z. 2D Tactile Sensor Based on Multimode Interference and Deep Learning. *Opt. Laser Technol.* **2021**, *136*, 106760. [CrossRef]
21. Deng, F.; Liu, Y.; Hou, Z.; Wang, Z.; Han, Z.; Hou, C.; Wang, M.; Yan, F. Optical Fiber Tactile Sensing Array Based on Dynamic Learning of Optical Specklegram. *Sens. Actuators A Phys.* **2026**, *402*, 117663. [CrossRef]
22. Huang, X.; Yan, J. Refractive Index Measurement Based on Multimode Fiber Specklegram Analysis Using a Deep Neural Network. *AIP Adv.* **2026**, *16*, 035205. [CrossRef]
23. Liu, Y.; Wang, J.; Fang, Q.; Zhang, K.; Li, Y.; Men, Y.; Yu, M.; Fan, H.; Lan, T.; You, J.; et al. A Low-Cost Fiber-Optic Temperature Sensor Utilizing Integrated Sensing and Leaky Mode Specklegram Analysis. *Opt. Lasers Eng.* **2026**, *196*, 109371. [CrossRef]
24. Wang, P.; Di, J. Deep Learning-Based Object Classification through Multimode Fiber via a CNN-Architecture SpeckleNet. *Appl. Opt.* **2018**, *57*, 8258–8263. [CrossRef] [PubMed]
25. Gomes, A.D.; Turtaev, S.; Du, Y.; Čižmár, T. Near Perfect Focusing through Multimode Fibres. *Opt. Express* **2022**, *30*, 10645. [CrossRef] [PubMed]
26. Borhani, N.; Kakkava, E.; Moser, C.; Psaltis, D. Learning to See through Multimode Fibers. *Optica* **2018**, *5*, 960. [CrossRef]
27. Resisi, S.; Popoff, S.M.; Bromberg, Y. Image Transmission Through a Dynamically Perturbed Multimode Fiber by Deep Learning. *Laser Photonics Rev.* **2021**, *15*, 2000553. [CrossRef]
28. Collard, L.; Kazemzadeh, M.; Piscopo, L.; De Vittorio, M.; Pisanello, F. Exploiting Holographically Encoded Variance to Transmit Labelled Images through a Multimode Optical Fiber. *Opt. Express* **2024**, *32*, 18896. [CrossRef] [PubMed]
29. Inalegwu, O.C.; Roman, M.; Mumtaz, F.; Zhang, B.; Nambisan, A.; Huang, J. Machine Learning-Assisted Optical Fiber Specklegram Sensor for Early and Spatially Distributed Water Leak Detection and Localization. *Opt. Eng.* **2025**, *64*, 056103. [CrossRef]
30. Yuan, H.; Liu, Y.; Li, G.; Pang, L.; Yang, Z.; Cai, Y.; Wang, M. Torsion and Bending Sensing Based on the Specklegrams from a Coupled Few-Mode Multi-Core Fiber. *Opt. Commun.* **2024**, *569*, 130732. [CrossRef]
31. Reja, M.I.; Smith, D.L.; Viet Nguyen, L.; Ebendorff-Heidepriem, H.; Warren-Smith, S.C. Multimode Optical Fiber Specklegram Pressure Sensor Using Deep Learning. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 7003910. [CrossRef]
32. Kang, J.; He, S.; Alsalman, O.; Liu, X.; Zhu, C. Deep Learning-Enabled Two-Directional Stretchable Strain Sensor Based on a Single Multimode Fiber. *IEEE Sens. J.* **2024**, *24*, 19143–19150. [CrossRef]
33. Wang, X.; Yang, Y.; Li, S.; Wang, X.; Zhang, P.; Lu, S.; Yu, D.; Zheng, Y.; Song, L.; Fang, F. A Reflective Multimode Fiber Vector Bending Sensor Based on Specklegram. *Opt. Laser Technol.* **2024**, *170*, 110235. [CrossRef]
34. Al Mahmud, R.; Iqbal, N.; Qureshi, K.K. Simultaneous Multi-Point Force Detection via Specklegram Analysis and Deep Learning. *IEEE Sens. J.* **2026**, *26*, 4004–4014. [CrossRef]

35. Arango, J.; Aristizabal, V.; Vélez, F.; Carrasquilla, J.; Gomez, J.; Quijano, J.; Herrera-Ramirez, J. Synthetic Dataset of Speckle Images for Fiber Optic Temperature Sensor. *Data Brief* **2023**, *48*, 109134. [CrossRef] [PubMed]
36. Vélez, F.J.; Arango, J.D.; Aristizábal, V.H.; Trujillo, C.; Herrera-Ramírez, J.A. Experimental Dataset for Fiber Optic Specklegram Sensing Under Thermal Conditions and Use in a Deep Learning Interrogation Scheme. *Data* **2025**, *10*, 44. [CrossRef]