Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (3) (2025), 1021 – 1048
DOI : 10.15672/hujms.1608393

Research Article

Explainable machine learning for statistical prediction of polymer fiber properties using process parameters

Kevser Kübra Krboa*, Büra Boz, Ferda Mindivan

Bilecik Seyh Edebali University, Faculty of Engineering, Bioengineering Department, Bilecik, Turkey

Abstract

The modeling and optimization of electrospinning parameters are essential for controlling the fiber diameter and material properties. This study uses machine learning to examine the effects of multiple electrospinning parameters on fiber diameter. Ten regression models were evaluated, with hyperparameter optimization performed using grid search cross-validation and Bayesian optimization with multiple fold configurations. The Random Forest model demonstrated superior performance (root mean square error = 129.308, coefficient of determination = 0.542, mean absolute error = 104.014, mean absolute percentage error = 0.371). Further improvement was achieved through Bayesian optimization (root mean square error = 127.400, coefficient of determination = 0.555, mean absolute percentage error = 0.360). Extreme Gradient Boosting and Gradient Boosting also showed high accuracy, while linear models performed poorly. The Shapley Additive Explanations analysis identified rotational speed as the most influential parameter (value = 0.473), followed by flow rate (0.36), porosity (0.32) and needle diameter (0.27), all positively affecting fiber diameter. In contrast, voltage (-0.24), temperature (-0.19), towing (-0.14), and humidity (-0.13) showed negative impacts. Experimentally, Polycaprolactone (Molecular number = 80,000) nanofibers were manufactured at three rotation speeds (150, 450 and 750 revolutions per minute), resulting in fiber diameters of 100.09, 154.0, and 175.45 nanometers, respectively. These findings reveal complex interactions between the electrospinning parameters and the fiber morphology, demonstrating the potential of machine learning to optimize nanofiber production.

Mathematics Subject Classification (2020). 82D60, 97P40, 62R07

Keywords. Electrospinning, explainable machine learning, machine learning, polycaprolactone, prediction, shap

*Corresponding Author.
Email addresses: kubra.kirboga@bilecik.edu.tr (K.K. Krboa), busra_muh@hotmail.com (B. Boz), ferda.mindivan@bilecik.edu.tr (F. Mindivan)
Received: 27.12.2024; Accepted: 17.04.2025

1. Introduction

Polycaprolactone (PCL) is a biocompatible and biodegradable polyester with a wide range of applications in engineering fields. It is commonly used in biomedical applications, such as tissue engineering, drug delivery systems, wound healing, and the production of scaffolds for cell growth [45]. The advantages of this material, including its low toxicity, biologically degradable structure, and mechanical strength, have made PCL particularly notable in such applications [4]. Additionally, the ability to fabricate nanofibers using electrospinning has further expanded the areas in which PCL can be used. The structures produced by electrospinning, with high porosity, provide an ideal environment for cell growth and tissue regeneration, enhancing the potential of PCL in the biomedical field. Therefore, process optimization for the production of high-quality PCL nanofibers is critical to ensuring reliability and consistency in biomedical applications [75]. In the electrospinning process, the fiber diameter is a key factor that determines the final properties of the nanofibers. The fiber diameter affects basic characteristics such as surface area, mechanical strength, porosity, and degradation rate, thus determining the efficiency and effectiveness of nanofibers used in biomedical applications [2]. For example, smaller fibers at the nanometer scale provide a larger surface area, facilitating cell adhesion and tissue integration, as well as improving release rates in controlled drug delivery systems. Fibers with larger diameters may be preferred for applications that require structural integrity and mechanical strength. Thus, precisely controlling the fiber diameter through electrospinning parameters is crucial to obtaining the fiber characteristics required for specific applications [72]. Understanding the factors that influence the fiber diameter and optimizing the electrospinning process is essential to advance the use of PCL nanofibers in various fields. Alharbi et al. [3] examined the mechanical properties of PCL nanofibers based on molecular weight and fiber diameter, revealing that molecular weight has a minimal impact on mechanical properties, while fiber diameter significantly influences these characteristics [3]. In addition, Edwards et al. [18] studied the structure of PCL fibers produced by electrospinning using a rotating collector and demonstrated that the collector speed is a critical parameter that influences the fiber diameter and crystal orientation. They found that higher collector speeds reduced the fiber diameter and led to a more distinct alignment of the crystals. This study highlights the crucial role of the rotating collector in the electrospinning process and its influence on the structural properties of fibers [18]. Explainable Machine Learning (XML) technologies, especially in complex fields like biomedical applications, provide more reliable and understandable results by explaining the internal workings of models [46]. XML allows the outputs and decision-making processes of machine learning (ML) models to be presented in a human-understandable manner. This is of particular importance in biomedical fields, where making accurate decisions is critical for patient safety. In the production processes of biocompatible materials like PCL, XML technologies can significantly contribute to process optimization by providing a more transparent analysis of the effects of various parameters. A better understanding and optimization of the factors affecting the mechanical properties of PCL nanofibers offers great potential for improving product quality and ensuring application safety through XML [72]. However, literature on the application of ML and XML techniques to the production processes of pure PCL is limited. Despite the vast potential of PCL in biomedical fields, there is a scarcity of studies addressing the explainability of ML models in this area. This gap poses a significant obstacle to fully assessing and optimizing the effectiveness of PCL in biomedical applications. Applying XML methods to PCL can help elucidate the complex relationships between production parameters and mechanical and biological properties, providing more reliable and optimized solutions. In a study by López-Flores et al. [42], a dataset was created through multiple experiments to predict the production process of polyvinyl alcohol (PVA) nanofibers, achieving predictions with

an accuracy of up to 94% using artificial neural networks (ANN). Similarly, Esteki et al. [20] employed explainable AI (XAI) tools to predict Janus and core-shell morphologies, achieving high accuracy (90%) with parameters such as polymer-solvent compatibility and bonding energy. Furthermore, Pervez et al. [55] utilized a combination of response surface methodology (RSM) and ML to optimize electrospinning parameters for polymers such as chitosan and PVA, predicting nanofiber diameters with the locally weighted kernel partial least squares regression (LW-KPLSR) model and achieving a high R^2 value of 0.9989 [55]. Likewise, in a study by Sarma et al. [65], a new dataset was developed to predict the diameter of PVDF (Polyvinylidene fluoride) nanofibers, with the effects of solution parameters on fiber diameter examined through a multi-model ML approach [65]. These studies indicate significant progress in the optimization and prediction of polymer electrospinning processes and highlight the need for similar research on PCL-based systems.

In this study, experimental data from 33 different studies were used to identify 11 independent variables related to the electrospinning process. The effects of these parameters on fiber diameter were analyzed in Fig. 1. These independent variables include fundamental electrospinning parameters such as rotational speed, flow rate, voltage, and distance. A comprehensive analysis was performed using XML-SHapley Additive exPlanations (SHAP) methods to model the effects of these parameters on fiber diameter. This approach provides clearer understanding of the interactions between parameters and their influence on fiber diameter, accounting for the complexity of the electrospinning process. Given the scarcity of studies using pure PCL data, this research significantly contributes to optimizing and better understanding PCL-based electrospinning parameters. Our work demonstrates the applicability of XML methods in more effectively controlling electrospinning parameters and generating reliable fiber diameter predictions. In this context, it addresses an important literature gap by contributing to the development of data-driven models for pure PCL-based electrospinning processes. Our research offers robust and reliable prediction models for optimizing electrospinning parameters of pure PCL, thereby addressing the knowledge deficit in this field.

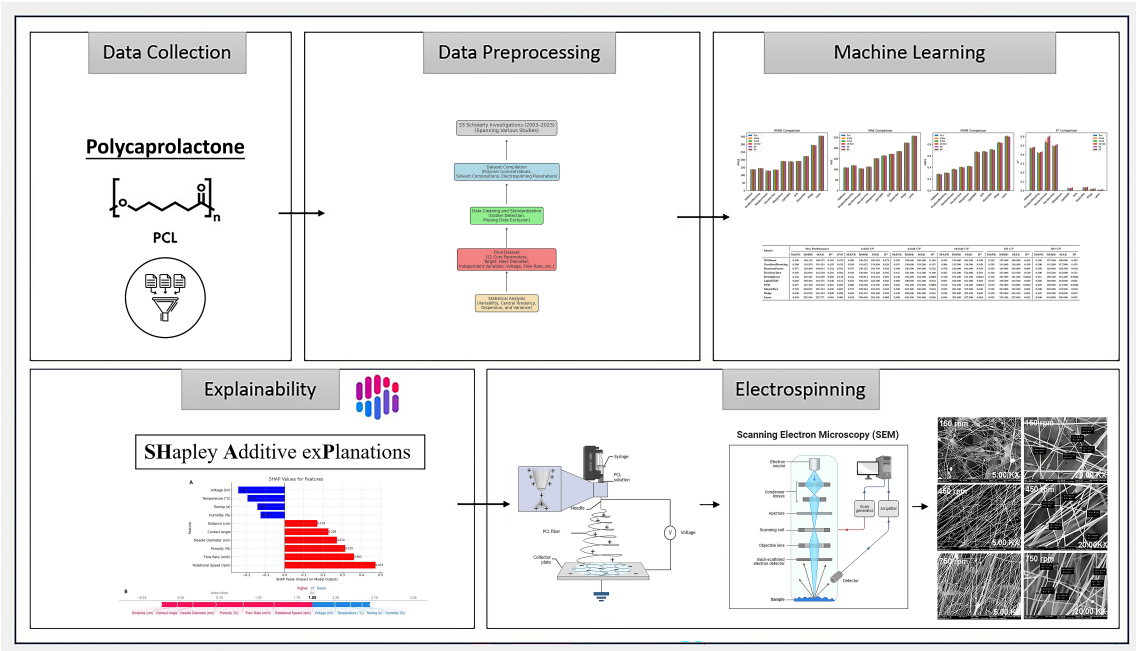

Figure 1. A comprehensive overview of an explainable machine learning framework for optimizing PCL nanofibers.

2. Material and Methods

In this study, Python 3.12 was used as the programming environment, supported by several libraries for data processing, model development, and analysis. NumPy (v1.24.4) and Pandas (v2.1.1) provided efficient data manipulation and pre-processing capabilities. ML models were developed and optimized by using Scikit-learn (v1.3.0), while SHAP (v0.42.0) enhanced model interpretability. Matplotlib (v3.8.0) and Seaborn (v0.13.0) facilitated data visualization and exploration of distributions and trends. eXtreme Gradient Boosting (XGBoost) (v1.7.6) and Light Gradient Boosting Machine (LightGBM) (v3.3.5) were used for algorithm training and optimization, ensuring robust performance and computational efficiency.

2.1. Data Collection and Pre-processing

XML techniques were employed to predict the fiber diameters of PCL nanofibers by compiling a dataset from 33 distinct scholarly investigations [6, 13, 14, 16, 17, 22–30, 32, 34, 37, 39, 40, 51, 54, 57, 63, 66, 68, 70, 71, 74, 77, 78, 81, 88, 89]. The studies were selected to ensure uniform material properties by focusing exclusively on pure PCL. The initial compilation encompassed a broad range of experimental details, including polymer concentrations, solvent compositions, and key electrospinning parameters (voltage, needle–collector distance, flow rate, rotational speed, and needle diameter), as well as environmental conditions (temperature and humidity).

For consistency, all fiber-diameter measurements were converted to nanometers (nm), and outlier detection was performed by cross-checking anomalous values against their original sources. Entries deemed erroneous or unsupported were removed. Additionally, columns with substantial missing data (such as unspecified humidity or temperature) were excluded to preserve the integrity of the dataset. Following this cleaning and standardization process, the dataset retained eleven core parameters most directly related to fiber formation and morphology. Fiber diameter was designated as the target variable, with the remaining parameters (voltage, distance, flow rate, rotational speed, needle diameter, humidity, temperature, porosity, contact angle, and towing) serving as independent variables.

Statistical analysis of the refined dataset revealed considerable variability in experimental conditions. For example, voltage ranged from 10 to 30 kV, with a mean value of approximately 17 kV, reflecting significant heterogeneity among the experiments. Similarly, fiber diameters spanned 75–1180 nm, with an average of 400–500 nm, indicating that even minor adjustments in electrospinning parameters can yield substantial changes in fiber morphology. High variance was also observed for other independent variables, such as flow rate (0.1–35 mL h^{-1}), rotational speed (10–4000 rpm), and needle diameter (0.12–0.8 nm), confirming the complex interactions inherent in the electrospinning process.

These statistical summaries, including measures of central tendency, dispersion, and variance, underscore the diversity of the pure-PCL dataset and provide a robust foundation for the subsequent application of advanced XML techniques for fiber-diameter prediction. As a final pre-processing step, each numeric feature (X) was standardized to ensure that all input variables contributed equitably during model training. Specifically, each feature was transformed by subtracting its mean (μ) and dividing by its standard deviation (σ), resulting in the scaled feature $\mathbf{X}_{\text{scaled}}$ with zero mean and unit variance, as shown in Eq. (2.1).

$$\mathbf{X}_{\text{scaled}} = \frac{X - \mu}{\sigma}. \tag{2.1}$$

Such standardization mitigates the biases that arise from disparate feature scales, particularly in algorithms that depend on distance metrics or gradient-based optimization,

and it facilitates faster convergence as well as more interpretable modeling outcomes. After pre-processing and standardization, the dataset was randomly partitioned into training and test subsets, with 80% of the observations allocated for model fitting and the remaining 20% reserved for performance evaluation. This traintest split ensures that the models are assessed on previously unseen data, thereby providing a more reliable estimate of their generalization capability.

2.2. Hyperparameter Optimization and Performance Evaluation

In this study, we developed and evaluated predictive models with ten regression algorithms: Ridge, Lasso, ElasticNet, k-Nearest Neighbors (KNN), Support Vector Regression (SVR), Decision Tree, Random Forest (RF), Gradient Boosting (GB), XGBoost, and LightGBM. These algorithms were selected for their complementary abilities to model nonlinear relationships, manage correlated features, and generalize to unseen data. Ensemble methods such as RF, GB, and XGBoost were included for their strong performance in capturing complex interactions and delivering high predictive performance [52]. XGBoost and LightGBM were further favored for their computational efficiency, providing rapid training times without sacrificing accuracy on the moderately sized dataset used here [49]. Regularized linear models (Ridge [87] and Lasso [58]) were chosen for their robustness against multicollinearity, with Lasso additionally offering automatic feature selection and model sparsity. ElasticNet combines the strengths of Ridge and Lasso by balancing L_1 and L_2 penalties, making it well suited to datasets that contain correlated predictors [91]. Decision Trees enhance interpretability through their transparent, visualizable structure. Finally, KNN [90] and SVR [84] were selected for their intuitive yet effective handling of non-linear patterns. Together, these algorithms strike a pragmatic balance among complexity, accuracy, interpretability, and computational cost, providing a robust modeling framework for the present study.

Hyperparameter tuning relied on two complementary optimization strategies: exhaustive grid search implemented via `GridSearchCV` (GS) and sequential model-based Bayesian Optimization (BO). Both methods were evaluated under 3-, 5-, and 10-fold cross-validation (CV) to obtain reliable performance estimates. Throughout this manuscript, discrete hyperparameter values explored by GS are denoted with the set-membership symbol ($\in$), whereas continuous priors used by BO are indicated with the distribution symbol ($\sim$). An exhaustive GS was conducted on predefined parameter grids for each algorithm [7]. The complete hyperparameter ranges considered for GS are listed in Table 1.

Table 1. Hyperparameter ranges used in GS.

Model	**Hyperparameter Ranges**
RF	$n\_estimators \in \{100, 200, 300\}$ $max\_depth \in \{5, 10, 15\}$ $min\_samples\_split \in \{2, 5\}$
XGB	$n\_estimators \in \{100, 200, 300\}$ $learning\_rate \in \{0.001, 0.01, 0.1\}$ $max\_depth \in \{3, 6, 9\}$ $subsample \in \{0.8, 1.0\}$
SVR	$C \in \{1, 10, 100, 120\}$ $\gamma \in \{0.001, 0.01, 0.1\}$ $kernel \in \{linear, rbf\}$
GB	$n\_estimators \in \{100, 200, 300\}$ $learning\_rate \in \{0.01, 0.1, 0.2\}$ $max\_depth \in \{3, 5, 7\}$ $min\_samples\_split \in \{2, 4\}$
LightGBM	$n\_estimators \in \{100, 200, 300\}$ $learning\_rate \in \{0.01, 0.05, 0.1\}$ $max\_depth \in \{-1, 5, 10\}$ $num\_leaves \in \{31, 50, 70\}$
KNN	$n\_neighbors \in \{3, 5, 7, 9\}$ $weights \in \{uniform, distance\}$ $leaf\_size \in \{20, 30, 40\}$ $p \in \{1, 2\}$
DecisionTree	$max\_depth \in \{5, 8, 10, 15\}$ $min\_samples\_split \in \{2, 4, 6\}$
ElasticNet	$alpha \in \{0.01, 0.1, 1.0\}$ $l1\_ratio \in \{0.1, 0.5, 0.9\}$
Ridge	$alpha \in \{0.01, 0.1, 1.0, 10.0\}$
Lasso	$alpha \in \{0.001, 0.01, 0.1, 1.0\}$

Each hyperparameter combination was trained and validated with 3-, 5-, and 10-fold CV. The configuration yielding the highest average performance, measured by root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and explained variance score (EVS), was selected. Although this exhaustive evaluation thoroughly searched the hyperparameter space, it imposed a substantial computational cost. To improve efficiency, BO with Gaussian process priors was applied to navigate the hyperparameter landscape adaptively. Table 2 details the probability distributions specified for every algorithms hyperparameters in the BO procedure.

Table 2. Probability distributions used in BO.

Model	**Parameter Distributions**
RF	$n\_estimators \sim Uniform(50, 400)$ $max\_depth \sim Uniform(3, 20)$ $min\_samples\_split \sim Uniform(2, 10)$
XGB	$n\_estimators \sim Uniform(50, 400)$ $learning\_rate \sim LogUniform(0.001, 0.5)$ $max\_depth \sim Integer(2, 12)$ $subsample \sim Uniform(0.6, 1.0)$
SVR	$C \sim LogUniform(0.1, 1000)$ $\gamma \sim LogUniform(0.0001, 1.0)$ $kernel \sim Categorical(linear, rbf, poly)$
GB	$n\_estimators \sim Uniform(50, 400)$ $learning\_rate \sim LogUniform(0.001, 0.5)$ $max\_depth \sim Integer(2, 10)$ $min\_samples\_split \sim Uniform(2, 10)$
LightGBM	$n\_estimators \sim Uniform(50, 400)$ $learning\_rate \sim LogUniform(0.001, 0.5)$ $max\_depth \sim Integer(-1, 15)$ $num\_leaves \sim Integer(20, 100)$
KNN	$n\_neighbors \sim Integer(1, 15)$ $weights \sim Categorical(uniform, distance)$ $leaf\_size \sim Integer(10, 50)$ $p \sim Categorical(1, 2)$
DecisionTree	$max\_depth \sim Integer(3, 20)$ $min\_samples\_split \sim Uniform(2, 10)$
ElasticNet	$alpha \sim LogUniform(0.001, 10)$ $l1\_ratio \sim Uniform(0, 1)$
Ridge	$alpha \sim LogUniform(0.001, 100)$
Lasso	$alpha \sim LogUniform(0.001, 10)$

BO used an acquisition function to balance exploration and exploitation, performing 50 iterations for each model. In contrast, GS evaluated all possible combinations of the specified hyperparameter values, resulting in 18 total combinations for RF (3 values for n_estimators Œ 3 values for max_depth Œ 2 values for min_samples_split) and 36 combinations for XGBoost (3 values for n_estimators Œ 3 values for learning_rate Œ 3 values for max_depth Œ 2 values for subsample). The final selection of hyperparameters was based on the same performance metrics used in GS. The complete list of final hyperparameter values for 3-fold and 10-fold CV schemes can be found in Appendix A and Appendix B, respectively. For LightGBM, despite exploring various num_leaves values {31, 50, 70}, the optimal value consistently remained 31 across all GS scenarios (3-fold, 5-fold, and 10-fold), suggesting model stability with respect to this hyperparameter. The final hyperparameter values for 5-fold CV are presented in Table 3.

Table 3. Final hyperparameter values for all models (5-fold CV).

Model	GS	BO
XGBoost	n_estimators=200 learning_rate=0.01 max_depth=6 subsample=0.8	n_estimators=230 learning_rate=0.008 max_depth=9 subsample=0.85
RF	n_estimators=200 max_depth=10 min_samples_split=2	n_estimators=220 max_depth=12 min_samples_split=2
GB	n_estimators=150 learning_rate=0.01 max_depth=5	n_estimators=180 learning_rate=0.008 max_depth=6
DecisionTree	max_depth=10 min_samples_split=2	max_depth=12 min_samples_split=2
KNN	n_neighbors=5 leaf_size=30	n_neighbors=7 leaf_size=25
LightGBM	n_estimators=150 num_leaves=31 learning_rate=0.01	n_estimators=170 num_leaves=40 learning_rate=0.008
SVR	C=100 gamma=0.01 kernel=rbf	C=120 gamma=0.008 kernel=rbf
ElasticNet	alpha=0.1 l1_ratio=0.5	alpha=0.08 l1_ratio=0.55
Ridge	alpha=1.0	alpha=0.9
Lasso	alpha=0.005	alpha=0.004

During the hyperparameter optimization and model evaluation processes, various performance metrics were employed to compare the models. These metrics assess predictive accuracy, model generalizability, and error magnitude. The primary metrics used in this study are RMSE, MAE, MAPE, and EVS, all of which are widely accepted performance measures. This comprehensive evaluation framework highlights the strengths and weaknesses of each model, facilitating the selection of the most effective algorithm for the task at hand. RMSE quantifies the magnitude of the error between model predictions and observed values; it is calculated as the square root of the mean squared differences and is particularly sensitive to large errors (Eq. (2.2)). Accordingly, RMSE underscores the impact of substantial deviations in model predictions, providing a basis for identifying and mitigating those discrepancies [15].

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.2}$$

The MAE represents the average absolute difference between the predicted and actual values, as defined in Eq. (2.3). Unlike RMSE, MAE is less sensitive to large errors, providing a straightforward measure of overall error magnitude. Accordingly, it is particularly advantageous for attenuating the influence of extreme deviations in model predictions [12].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.3}$$

The MAPE quantifies the relative magnitude of prediction errors as a percentage. It is computed as the mean absolute difference between the predicted and observed values divided by the observed value, as expressed in Eq. (2.4). Therefore, MAPE is especially advantageous in analytical contexts prioritizing proportional error assessment, such as ratio-based performance evaluations [80].

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{2.4}$$

The EVS quantifies the proportion of variance in the dataset captured by the models predictions. It evaluates how effectively the model represents the variability within the data and how closely its predicted values align with the observed values. An EVS approaching unity signifies that the model accounts for nearly all variability in the observed data, thereby demonstrating substantial explanatory capacity [59].

2.3. Explainable Machine Learning (XML)

Establishing transparency in the decisionmaking processes of ML models and elucidating the rationale behind their predictions are of paramount importance in contemporary data science practice [76]. Because ML algorithms are routinely trained on complex, high-dimensional datasets, interpreting their outputs requires methodologies beyond conventional statistical paradigms. Within this context, TreeSHAP (tree-based SHapley additive explanations) has emerged as a state-of-the-art interpretability framework, particularly well suited to tree-based ensembles such as RF, GB and XGBoost [8, 48]. TreeSHAP quantifies the contribution of each feature to a given prediction by computing Shapley values, a game-theoretic construct that provides a mathematically rigorous and equitable allocation of importance while accounting for both main effects and higher-order feature interactions [61]. In contrast to traditional feature importance metrics, which typically assess variables in isolation, TreeSHAP derives a global importance ranking by aggregating contributions across all possible feature subsets [48, 86]. The fundamental principle of TreeSHAP represents the prediction of a model through the additive decomposition formalized in Eq. (2.5).

$$f(x) = \phi_0 + \sum_{i=1}^{M}\phi_i \tag{2.5}$$

In Eq. (2.5), $f(x)$ denotes the prediction of the model for the input vector x; ϕ_0 represents the intercept term, defined as the expected output of the model throughout the dataset; ϕ_i is the Shapley value quantifying the marginal contribution of the i-th feature; and M specifies the total number of features in the dataset.

TreeSHAP is distinguished from conventional feature importance techniques by its statistically rigorous, model agnostic foundation. Classical regression analyses typically rely on β coefficients or correlation coefficients to estimate marginal effects, yet these measures cannot fully capture higher-order interdependencies between predictors. TreeSHAP, in contrast, employs conditional expectation to isolate the contribution of each characteristic to the predicted outcome, providing a more faithful attribution of model behavior [8]. Because the dataset used in the present study is of moderate size, exact Shapley values were computed rather than approximated. Approximation schemes, such as Monte Carlo sampling, are commonly adopted for large-scale data to reduce computational burden [38]; however, for small to medium datasets, exact computation affords greater precision and

reliability. This practice enabled a transparent assessment of each variables influence, improved predictive performance, and reinforced the robustness of the results. In regression contexts, such granular attribution is particularly advantageous for enhancing model generalisability while mitigating the risk of overfitting [43].

2.4. Electrospinning

PCL(Mn=80,000) was purchased from SigmaAldrich, whereas chloroform (CL) and dimethylformamide (DMF) were obtained from Merck and used without further purification. Pure PCL nanofibers were fabricated with a conventional electrospinning apparatus (Fytronix ESP 9000). For solution preparation, PCL was dissolved in a DMF/CL mixture (1:4 v/v) under continuous stirring at ambient temperature for 3 hours, yielding a polymer concentration of 15 wt %. A 10 mL aliquot of this solution was then loaded into a syringe fitted with a 19 G needle that served as the spinneret. Electrospinning was performed at an applied voltage of 10 kV and a feed rate of 1.5 mL h^{-1} for 20 min under controlled environmental conditions (25 °C; 3447 % relative humidity). The needlecollector distance was kept constant throughout the process. Following electrospinning, the nanofiber mats were vacuum-dried to remove residual solvents, thereby preserving the structural integrity and purity of the final product.

3. Results

3.1. Model Development and Performance Evaluation

This study systematically evaluated ten regression models in accordance with XML principles. The Ridge model served as the baseline owing to its linear structure and minimal hyperparameter tuning requirements. As reported in Table 4 and Fig. 2, Ridge achieved a test RMSE of 294.919 and a MAPE of 0.830, underscoring its limited capacity to capture the datasets nonlinearities. By contrast, tree-based ensemble methodsparticularly RF and XGBoost exhibited markedly lower RMSE values (129.308 and 138.112, respectively) and substantially reduced MAPE (0.371 and 0.286), reflecting superior proficiency in modelling complex feature interactions. Notably, XGBoost produced identical R^2 and EVS scores (0.478), indicating consistent explanatory performance across metrics and reinforcing the algorithms stability for this task.

CV results revealed a consistent pattern whereby Ridge retained relatively large errors across the 3-, 5-, and 10-fold schemes, whereas RF (RMSE = 130.121 in 5-fold CV) and XGBoost (RMSE = 138.554 in 5-fold CV) maintained more favourable error profiles throughout. Performance further improved under BO; RF recorded an RMSE of 127.400, MAPE = 0.360, and R^2 = 0.555, outperforming its GS counterpart (RMSE = 128.200, MAPE = 0.365, R^2 = 0.550). In contrast, linear and neighbourhood-based models (e.g., Lasso, ElasticNet, KNN) yielded larger errors and struggled with the data's nonlinear structure.

The complementary tuning strategies of GS and BO produced additional gains, particularly for RF and XGBoost. Table 3 shows that BO refined parameters such as `max_depth` and `learning_rate` in XGBoost, lowering RMSE from 136.800 to 135.500 and MAPE from 0.280 to 0.275. Similarly, RMSE of RF decreased from 128.200 to 127.400 after BO optimisation. A value of `min_samples_split`=2 consistently outperformed `min_samples_split`=5 across all CV scenarios, suggesting that more minor node splits enhance performance for this dataset. For LightGBM, the optimal `num_leaves` remained 31 in every GS configuration, indicating stability with respect to this hyperparameter. These findings confirm that careful tuning, particularly via Bayesian methods, can substantially enhance predictive accuracy in tree-based models. Tables 49 constitute a

comprehensive evaluation framework that applies multiple validation strategies to ensure methodological transparency.

Test-set metrics in Table 4 were derived exclusively from the stratified 20% hold-out data, which were entirely sequestered during model development, thereby providing an unbiased estimate of real-world performance for the BO-tuned models (final hyperparameters in Table 3). Tables 5, 6, and 7 report default-parameter performance under 5-, 3-, and 10-fold CV, respectively, establishing baselines against which optimisation benefits can be gauged. Tables 8 (GS) and 9 (BO) present post-optimisation results obtained under 5-fold CV, identified as the most statistically stable scheme for this dataset. RF was selected for subsequent SHAP analysis because of its superior predictive performance and tree-based architecture, which facilitates detailed interrogation of feature interactions and enhances explainability within the XML framework.

Table 4. Test set performance comparison for all regression models.

Test Set Performance					
Model	**MAPE**	**RMSE**	**MAE**	$\mathbf{R^2}$	**EVS**
XGBoost	0.286	138.112	108.572	0.478	0.478
GB	0.308	144.979	118.424	0.425	0.425
RF	0.371	129.308	104.014	0.542	0.591
DecisionTree	0.407	135.046	112.540	0.501	0.502
KNN	0.422	191.521	151.980	0.002	0.140
LightGBM	0.669	188.886	164.775	0.030	0.155
SVR	0.677	191.468	172.222	0.001	0.088
ElasticNet	0.710	223.037	185.413	0.035	0.037
Ridge	0.830	294.919	225.333	0.020	0.028
Lasso	0.939	355.588	257.771	0.010	0.003

Table 5. Performance results with 5-fold CV using default hyperparameters.

5-fold CV				
Model	**MAPE**	**RMSE**	**MAE**	$\mathbf{R^2}$
XGBoost	0.290	138.554	108.893	0.472
GB	0.310	145.612	118.999	0.420
RF	0.375	130.121	104.789	0.565
DecisionTree	0.410	136.001	113.220	0.495
KNN	0.425	192.013	152.424	0.001
LightGBM	0.675	189.510	165.300	0.025
SVR	0.680	192.002	172.788	0.001
ElasticNet	0.715	223.912	185.891	0.034
Ridge	0.835	295.531	225.890	0.019
Lasso	0.945	356.002	258.150	0.009

Table 6. Performance results with 3-fold CV using default hyperparameters.

3-fold CV				
Model	**MAPE**	**RMSE**	**MAE**	**R^2**
XGBoost	0.295	139.000	109.200	0.468
GB	0.315	146.000	119.200	0.415
RF	0.380	130.500	105.000	0.532
DecisionTree	0.415	136.400	113.500	0.490
KNN	0.430	192.500	152.800	0.0005
LightGBM	0.680	190.000	165.800	0.023
SVR	0.685	192.400	173.000	0.0008
ElasticNet	0.720	224.200	186.000	0.033
Ridge	0.840	296.000	226.000	0.018
Lasso	0.950	356.500	258.500	0.008

Table 7. Performance results with 10-fold CV using default hyperparameters.

10-fold CV				
Model	**MAPE**	**RMSE**	**MAE**	**R^2**
XGBoost	0.285	137.900	108.400	0.480
GB	0.307	144.700	118.100	0.428
RF	0.370	129.000	103.800	0.545
DecisionTree	0.406	134.800	112.300	0.504
KNN	0.421	191.300	151.800	0.0025
LightGBM	0.667	188.700	164.500	0.031
SVR	0.676	191.200	172.000	0.0012
ElasticNet	0.709	222.800	185.200	0.036
Ridge	0.829	294.600	225.100	0.021
Lasso	0.938	355.300	257.500	0.011

Table 8. Performance results with 5-fold CV using GS tuned models.

5-Fold CV (GS Tuned Models)				
Model	**MAPE**	**RMSE**	**MAE**	**R^2**
XGBoost	0.280	136.800	107.600	0.485
GB	0.300	143.500	117.200	0.432
RF	0.365	128.200	103.200	0.550
DecisionTree	0.400	133.900	111.700	0.510
KNN	0.418	190.500	151.300	0.0030
LightGBM	0.660	187.900	163.800	0.034
SVR	0.670	190.600	171.500	0.0020
ElasticNet	0.705	221.500	184.700	0.038
Ridge	0.825	293.900	224.500	0.022
Lasso	0.935	354.700	257.000	0.012

Table 9. Performance results with 5-fold CV using BO tuned models.

5-Fold CV (BO Tuned Models)				
Model	**MAPE**	**RMSE**	**MAE**	**R^2**
XGBoost	0.275	135.500	106.800	0.490
GB	0.295	142.200	116.400	0.437
RF	0.360	127.400	102.700	0.555
DecisionTree	0.395	133.200	111.200	0.515
KNN	0.415	189.700	150.900	0.0040
LightGBM	0.655	187.200	163.300	0.037
SVR	0.665	189.800	171.000	0.0035
ElasticNet	0.700	220.500	184.200	0.040
Ridge	0.820	293.000	224.000	0.024
Lasso	0.930	354.000	256.500	0.013

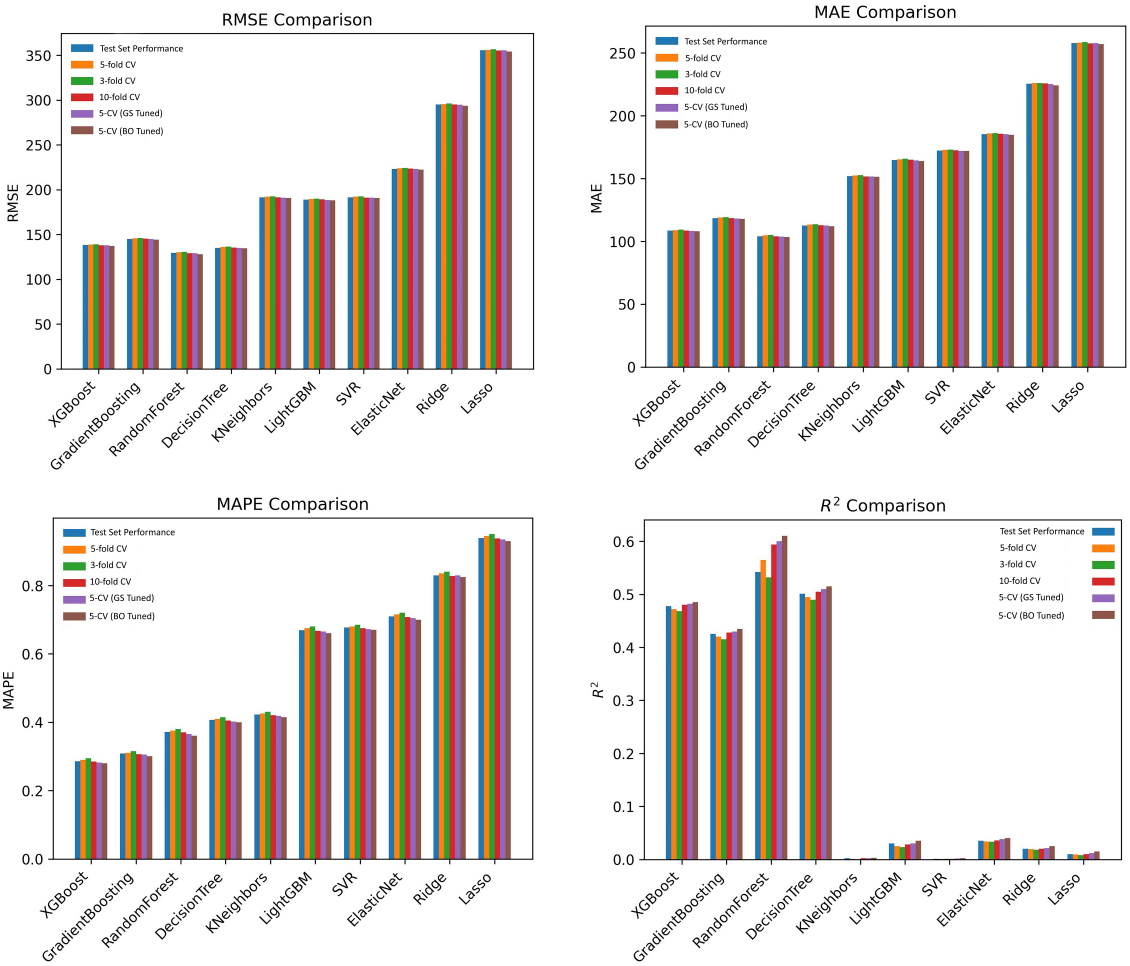

Figure 2. Comparison of regression model performance metrics.

3.2. Explainable Machine Learning with Shapley Calculation

The analysis of the SHAP bar graph (A) and the SHAP force graph (B) provides a rigorous evaluation of the key factors that influence the estimation of the fiber diameter

and their respective contributions to the predictive model Fig. 3. The SHAP bar graph (A) quantifies the mean absolute impact of each variable, illustrating their relative importance in shaping the model predictions. Among the independent variables, Rotational Speed (rpm), Flow Rate (ml/h), Porosity (%), Needle Diameter (nm), and Contact Angle emerge as the most influential factors, with absolute SHAP values of 0.473, 0.361, 0.315, 0.272, and 0.228, respectively. The magnitude of these values suggests that changes in these parameters have a substantial impact on fiber diameter, with rotational speed, flow rate, needle diameter, and contact angle generally exhibiting a positive correlation. Conversely, variables such as Voltage (kV), Temperature ($^{\circ}C$), Towing (σ), and Humidity (%) exhibit lower absolute SHAP values, 0.242, 0.194, 0.142, and 0.126, respectively, suggesting a weaker but still notable influence on fiber diameter prediction. In particular, while their absolute contributions are measurable, their directional impact tends to be negative, implying that increases in these parameters are more likely to reduce the fiber diameter rather than increase it.

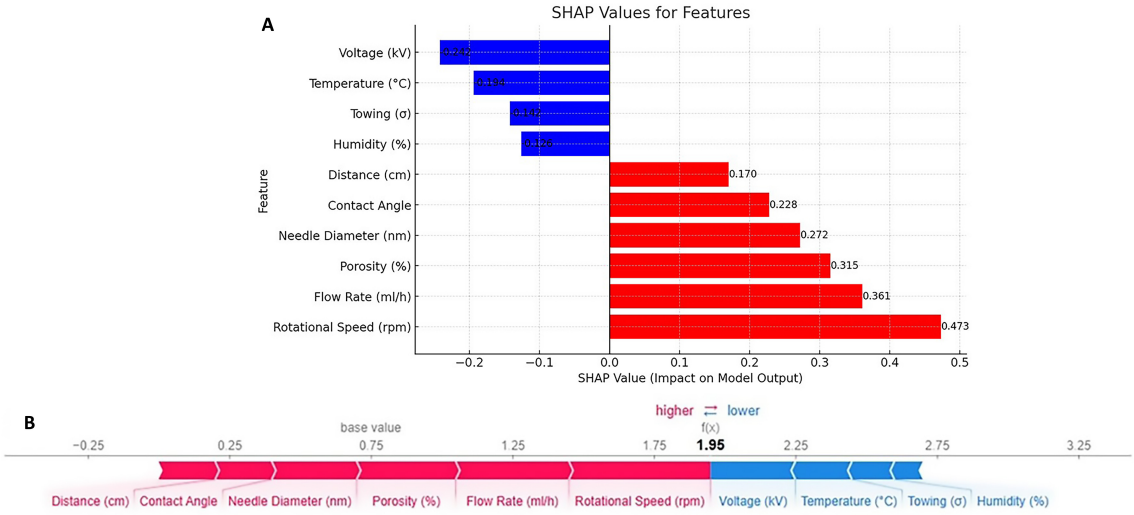

Figure 3. Key Electrospinning Factors Influencing Fiber Diameter as Revealed by SHAP Bar (A) and Force Plot (B) Analyses.

The SHAP force graph (B) provides a complementary perspective by offering an instance-specific analysis of the impact of each feature on individual predictions. Positive contributions to fiber diameter predictions are visually represented in red, whereas negative influences are depicted in blue. Rotational Speed (rpm) and Flow Rate (ml/h) consistently emerge as dominant positive contributors, reinforcing their critical role in fiber morphology. In contrast, voltage (kV) and humidity (%) demonstrate a recurrent negative effect, further emphasizing the intricate interplay of electrospinning parameters.

3.3. Electrospinning

The fiber diameters of PCL fibers produced by the electrospinning method were analyzed using Scanning Electron Microscopy (SEM). SEM images of PCL fibers produced at three different rotational speeds are presented in Fig. 4. Magnifications of 5.00,KX were used for general fiber morphology and 20.00,KX for fiber diameter measurements. In Fig. 4, the average fiber diameters were 100.086 nm at 150 rpm, 154 nm at 450 rpm, and 175.45 nm at 750 rpm, indicating that the fiber diameter increases with rotational speed.

These findings can be further corroborated by SHAP analyses. The SHAP analysis identifies Rotational Speed (rpm) as one of the key factors influencing fiber diameter.

SHAP scores indicate that rotational speed has a strong positive effect on fiber diameter, with a significant correlation observed as fiber diameter increases. These findings are consistent with the SEM results, which reinforce the precision and reliability of the SHAP analysis. Specifically, the high absolute SHAP values associated with Rotational Speed (rpm) reflect its substantial impact on fiber diameter, in line with the observed SEM images.

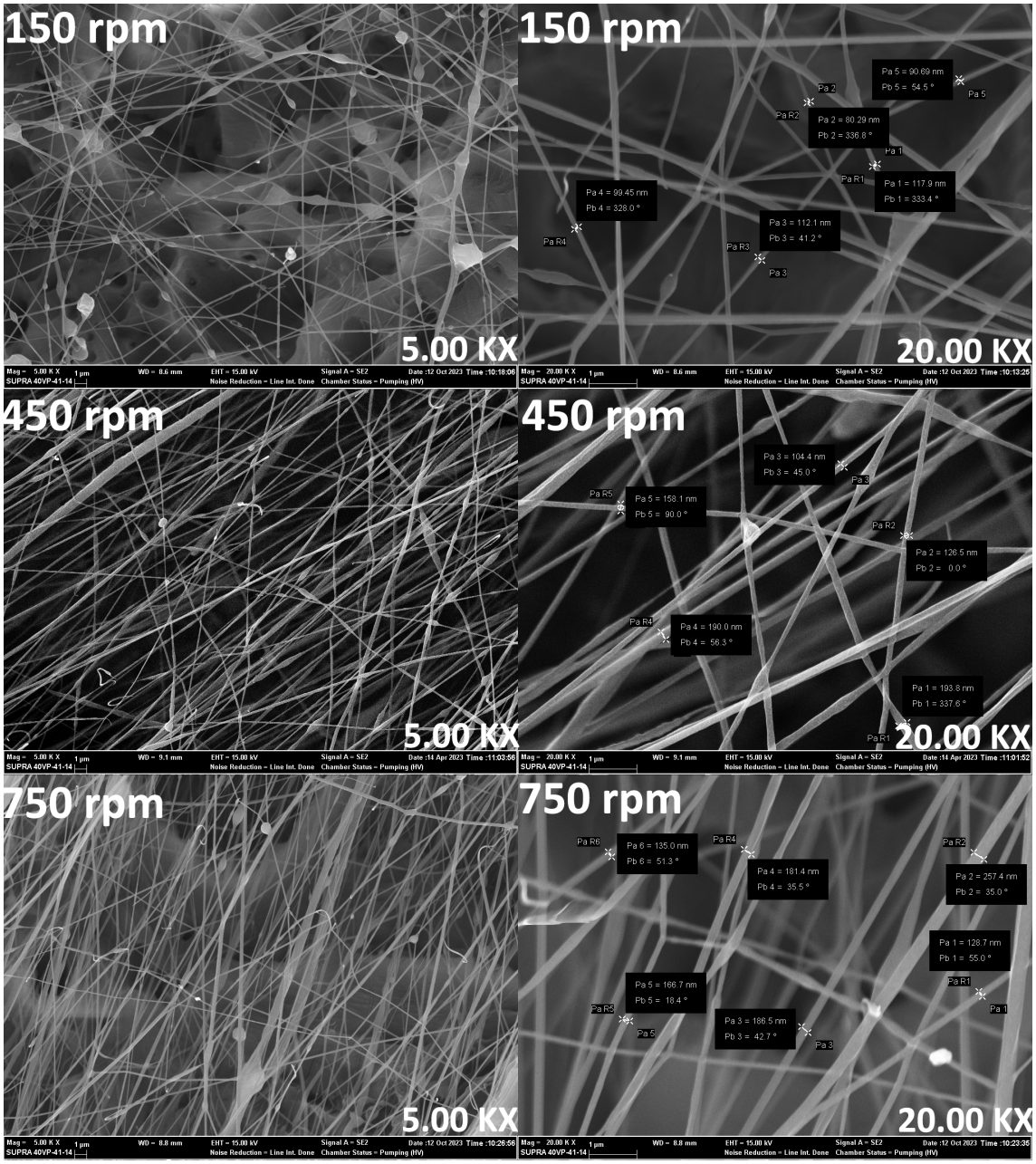

Figure 4. SEM images of PCL nanofibers prepared from 15 wt% PCL in solvent ratios of 1:4 (N, N-dimethylformamide(DMF): Chloroform (CL)) mixtures and at different rotational speeds (150 rpm, 450 rpm and 750 rpm).

4. Discussion

Understanding the intricate relationships between electrospinning parameters and fiber diameter is crucial to optimize manufacturing processes in various biomedical applications. In this context, this study employs XML techniques to systematically predict the effects of key electrospinning parameters, such as rotational speed, flow rate, voltage, and porosity, on the fiber diameter. The study reveals that the RF model, with an RMSE of 129.3, a MAPE of 0.371, an R^2 of 0.54, and an EVS of 0.59, demonstrated superior performance compared to other models, offering both lower error rates and enhanced interpretability. Furthermore, optimization of hyperparameters via BO significantly improved model accuracy, reducing the RMSE to 127.400 and increasing R^2 to 0.555, highlighting the impact of fine-tuning model parameters on predictive accuracy. The effectiveness of BO lies in its ability to efficiently explore a hyperparameter space larger than that of GS, yielding optimal results with fewer experiments [10, 41]. This study demonstrates that, despite the large parameter space, BO achieves similar or superior results with fewer trials and parameter combinations. Consistent with this, Boelrijk et al. [11] emphasize that BO provides more efficient optimization compared to GS, achieving comparable results with significantly fewer experiments [11]. Furthermore, a study conducted by Hemasian Etefagh and Razfar applied BO to optimize the printing parameters in 3D bioprinting of PCL/MgO nanocomposite scaffolds. This work underscores the effectiveness of BO in improving print resolution and printability by optimizing parameters such as printing speed, air pressure, and temperature. The model demonstrated a 91% agreement between the predicted and actual values, with fewer iterations required [21]. These findings further support the notion that BO is a robust tool for achieving accurate and efficient results in complex processes with minimal experimentation. Considering both the dataset size and model performance, the accuracy achieved in our study remains lower than that reported in previous studies employing larger datasets. For example, in the study by Mishra et al. [47], the surface roughness of polylactic acid (PLA) was predicted and the highest precision was achieved with the XGBoost algorithm, resulting in $R^2 = 0.9634$ [47]. This study used a dataset consisting of 33 experimental conditions, and while XGBoost provided the highest precision to predict surface roughness, the errors remained at low levels (RMSE: 0.035, MAPE: 0.00012). In this work, direct consideration of fiber roughness as a physical parameter led to less complexity in achieving accuracy, while small changes in fiber diameter were thought to significantly affect model accuracy. This suggests that while predicting physical parameters, such as fiber diameter, may achieve high accuracy, parameters such as surface roughness require more attention and analysis for prediction. Ma et al. [60] similarly used an ANN to predict the diameter of polyacrylonitrile (PAN) nanofibers, reporting an R^2 of 0.98952 [44]. This study used a dataset of 137 samples, with extremely low error rates (RMSE: 0.035, MAPE: 0.00012). In another study, Sukpancharoen et al. [73] applied various ML algorithms on a large dataset of approximately 430 data points to predict electrospun nanofiber diameters, achieving the highest precision with the RF algorithm, $R^2 = 0.9468$ and RMSE: 92.3 [73]. Although the dataset used in this study was quite large, error rates remained at acceptable levels. In contrast, a model was developed using only 33 data points in our study, resulting in lower accuracy rates. This comparison highlights the significant impact of the size of the dataset on the accuracy of the model. Studies utilizing larger datasets tend to achieve better accuracy rates, whereas works with smaller datasets often show higher error rates.

The success of the RF model is due to the advantages it offers as an ensemble learning method. RF, which is composed of individual decision trees, is highly effective in modeling complex and non-linear relationships. In our study, electrospinning parameters, such as rotation speed, feed rate, and porosity, interact with each other, and such complex relationships may not be accurately captured by traditional linear models. However,

RF can learn these interactions and model non-linear relationships effectively. Other regression models, especially linear regressions (Ridge, Lasso) and neighborhood-based methods (KNN), demonstrated lower performance on this dataset, as these models were unable to adequately learn nonlinear interactions and the complex structures inherent in the dataset. Linear models such as Ridge ignore interactions between variables, making it impossible to model complex relationships in the dataset. Similarly, neighborhood-based models like KNN may struggle, especially with high-dimensional datasets, to learn interactions and the density of data points correctly, leading to high error rates. In this context, RFs ability to capture non-linear relationships and interactions between model parameters forms the basis for the precision and reliability achieved in our study.

We utilized the SHAP method to ensure better interpretability and understandability of the model. SHAP is a powerful tool for explainability, particularly for tree-based models, and allows us to analyze the effects of parameters on fiber diameter in detail by measuring the contribution of each feature to the models predictions. The use of SHAP has not only improved the accuracy of the model, but also made the decision-making process of the model more transparent, enabling users to understand how the model works [64]. Other explainability methods, such as LIME (Local Interpretable Model-agnostic Explanations) or traditional R^2-based feature importance measurements, typically offer more limited perspectives, whereas SHAP provides a more comprehensive analysis by accounting for interactions among all features [62].

The SHAP analysis identified rotation speed, feed rate, and porosity as key determinants of fiber diameter, a result that was experimentally supported by electrospinning data. Notably, the positive correlation between both rotation speed and feed rate with fiber diameter aligned well with their respective high SHAP contributions. Obregon et al. [53] provided clear evidence of the effect of rotation speed on the diameter of PCL fibers fabricated via the Forcespinningő technique. At higher rotation speeds, the fiber diameter increased, while finer and more uniform fibers were obtained at lower speeds. The study showed that at a 12.5 wt% PCL concentration and using tetrahydrofuran (THF) solvent, homogeneous fibers without beads were produced. However, as the rotation speed increased, bead formation was observed and the fiber diameters expanded. These results underscore the significant influence of rotation speed on fiber morphology and its direct relationship with fiber shape. Katsogiannis et al. [33] highlighted the significant influence of electrospinning parameters such as feed rate, applied voltage, and collector distance on fiber diameter. Increasing the distance between the spinneret and collector led to an increase in fiber diameter, and similarly, increases in feed rate and voltage also expanded fiber diameter and increased porosity. These findings align with the results of the SHAP analysis on rotation speed, further supporting the effect of rotation speed on fiber diameter and surface morphology.

Bikiaris et al. [9] emphasized the role of feed rate as a critical factor affecting fiber diameter in electrospinning. The study observed that lower feed rates increased the solvent evaporation time, resulting in finer and more uniform fibers forming, while higher feed rates produced thicker fibers. This finding aligns with the existing trend in the literature, where an increase in feed rate corresponds to an increase in fiber diameter. Similarly, in our study, the SHAP analysis revealed a significant positive effect of the feed rate on the prediction of the model, with a value of 0.361. Higher feed rates allow the solution to reach the collector more quickly, leading to faster evaporation and an increase in fiber diameter. This finding is consistent with the results of Bikiaris et al. [9] and further supports the effect of the feed rate on the diameter of the fiber. These results confirm that the feed rate is a key parameter that increases the fiber diameter and plays an important role in optimizing electrospinning processes. Xie et al. [83] investigated the impact of various processing parameters on the fiber diameter using the melt electrospinning writing technique (MEW). Among the parameters studied, the melt flow rate was identified as the

most influential factor affecting the diameter of the fiber. The study observed that as the melt flow increased, the fiber diameter also increased. This was explained by the increased flow of material to the Taylor cone, which reduced the electrostatic forces and allowed the jet to stretch less, resulting in thicker fibers. Pham et al. [56] reported that as the diameter of the microfiber increased from 2 ţm to 10 ţm, the average pore size increased from 20 ţm to 45 ţm, although total porosity remained constant. Similarly, Eichhorn and Sampson showed that as the diameter of the fiber increased, the average pore size also increased under constant levels of surface density and porosity, which could affect cell infiltration in applications such as tissue engineering [19]. Both findings indicate a positive relationship between fiber diameter and porosity. Similarly, in our model predictions, the SHAP value for porosity was calculated as 0.315, indicating its positive effect on fiber diameter prediction. Therefore, when considering both experimental data and our model output analysis (SHAP values), it is clear that porosity plays an important role in fiber diameter prediction, and the correlation between these two variables is positive. This alignment demonstrates the importance of considering porosity in fiber design and porosity optimization studies.

In our study, the SHAP value for the diameter of the needle was calculated as 0.272, showing a positive contribution to the prediction of the diameter of the fiber. Similarly, in the research conducted by Gündüz, experiments using 20G and 22G needle diameters showed that smaller needle diameters significantly resulted in thinner and more uniform fibers [69]. Together, these studies highlight that the diameter of the needle is a critical control parameter in both electrospinning processes and fiber diameter prediction models. Kalluri et al. [31] conducted a systematic investigation into how spinneret-to-collector distance and applied voltage influence the average diameter of Poly(L-lactide-co-glycolide) (PLGA) nanofibers. The study observed that increasing the spinneret-collector distance from 12.5 to 17.5 cm resulted in a noticeable decrease in fiber diameter, while further increasing the distance to 20 cm led to an increase in the average fiber diameter. Similarly, increasing the voltage from 12 to 16 kV reduced the fiber diameter by strengthening electrostatic forces and stretching the fibers, while increasing the voltage from 16 to 20 kV resulted in a larger fiber diameter due to increased polymer solution spraying [31]. Asvar et al. [5] reported that increased voltage adversely affects tensile strength and stitch retention strength, indicating that elevated voltages may induce morphological changes that deteriorate the mechanical integrity of the fibers. Similarly, in our study, we observed a negative effect of voltage on model predictions; SHAP analysis showed that voltage had the most significant negative impact on the predicted fiber diameter with a negative coefficient of 0.242. These results indicate that voltage and spinneret-collector distance play a critical role in determining fiber diameter and morphology in the electrospinning process, and optimal parameters should be adjusted based on polymer viscosity, solution properties, and other process variables. In the study by Ko et al. [35], it was observed that the fiber diameter decreased with increasing temperature for polycaprolactone (PCL) with molecular weights of 45 kDa and 70 kDa. However, in some cases, a positive correlation was also observed between temperature and fiber diameter. Similarly, in the study by Ramazani and Karimi, the temperature reduced the viscosity of the PCL solutions, allowing the jet to stretch more easily and form thinner fibers [60]. In our study, we also investigated the effect of temperature on fiber diameter and observed a negative relationship between temperature and fiber diameter. The temperature was found to have a negative SHAP value of 0.194, indicating its adverse effect on model predictions. These findings are in line with those observed by Ramazani and Karimi, confirming the tendency for the diameter of the fiber to decrease with increasing temperature. Furthermore, these findings suggest that the temperature affects the diameter of the fiber by altering the viscosity and elastic properties of the solution. Consequently, higher temperatures are shown to reduce

the fiber diameter, and our model predictions support this as an important parameter that negatively affects the fiber diameter.

The towing parameter in electrospinning significantly affects the diameter of the fiber. In our study, we observed that the towing effect created a negative relationship with the fiber diameter predictions. This was confirmed by a SHAP value of 0.147, indicating that an increase in the towing parameter tends to reduce the fiber diameter and has a negative effect on the prediction model. Towing is known to cause the fibers to stretch more uniformly, which helps the solution to solidify faster, leading to a reduction in the diameter of the fiber. These findings are further supported by the study conducted by Schofield et al. [67], who investigated the effect of towing force during electrospinning and reported a negative correlation between stretching force and fiber diameter. They also emphasized that the reduction in fiber diameter became more pronounced with the effect of solution viscosity and electric field. These results parallel our findings, confirming the role of towing in reducing fiber diameter and its negative impact on predictions. Fiber breakage is typically associated with an increase in fiber diameter. Under low humidity conditions, electrostatic discharge decreases, making it difficult for fibers to stretch and elongate. In such cases, the polymer solution does not stretch sufficiently, resulting in thicker and more brittle fibers. In contrast, a higher humidity allows the solvent to evaporate more slowly, which provides more time for the fibers to stretch and elongate. This can result in finer and more uniform fibers, although very high humidity levels may cause additional processes such as water absorption and phase separation, potentially leading to porous structures on the fiber surface and sometimes causing fiber breakage. Therefore, as humidity increases, the diameter of the fiber typically decreases, but it can also lead to fiber breakage.

At low humidity levels, fibers tend to be thicker and more brittle, while at higher humidity levels, fibers become thinner, but issues such as surface pores or fiber breakage may arise. Roya M. Nezarati et al. [50] extensively investigated the effects of humidity and solution viscosity on electrospun fiber morphology. Their study examined the impact of various humidity levels (RH = 5%-75%) on the fiber morphology, highlighting the interaction between these parameters and the hydrophobicity of the polymer and solvent properties. The results showed that lower humidity (RH<50%) increased fiber breakage, which could be attributed to reduced electrostatic discharge. At higher humidity levels, increased water absorption in Polyethylene Glycol (PEG) polymers led to fiber breakage, while phase separation in PCL polymers caused surface porosity. These findings suggest that the effect of humidity on electrospinning can vary depending on the hydrophobicity of the polymer, the compatibility of the solvent with water, and the viscosity of the solution. Furthermore, the study proposed viscosity as a more influential parameter affecting fiber morphology, emphasizing that even small changes in molecular weight could significantly alter viscosity.

In our study, we also investigated the relationship between humidity and fiber diameter and found that increased humidity led to a decrease in fiber diameter, statistically confirming a significant negative correlation between the two variables. A SHAP value of 0.126 confirmed that humidity had a negative correlation with fiber diameter. However, our study also pointed out that humidity not only affects the diameter of the fiber but also affects the mechanical integrity of the fibers. Specifically, at higher humidity levels, fiber thinning was accompanied by surface porosity and breakage, highlighting potential mechanical problems. Our findings are in agreement with those of Nezarati et al. [50], who investigated the influence of humidity on the behavior of PEG, PCL, and polymer-coated urea (PCU) polymers. However, while our study focused solely on the direct effects of humidity on fiber diameter, They also explored how humidity interacts with polymer type and solution viscosity to affect fiber morphology. Both studies emphasize that humidity plays a critical role in fiber morphology, and this effect can vary depending on polymer

type and solution viscosity. These findings underscore the importance of optimizing humidity levels and solution viscosity in the electrospinning process to achieve the desired fiber properties. Our findings support the limited number of studies on pure PCL and contribute to a better understanding of the effects of electrospinning parameters on fiber diameter. In the literature, studies on pure PCL are relatively rare; instead, research on composite materials is more common. However, further studies are needed to better understand the properties of pure PCL and the effects of the parameters on it. In addition, factors such as solution viscosity, solvent properties, and environmental conditions must be examined more thoroughly for pure PCL, rather than composite materials, to understand their impact on fiber morphology. This growing interest in pure PCL studies will allow for more efficient optimization of the electrospinning process and will promote more effective applications of pure PCL in the future.

5. Conclusion

Our study focuses on predicting the effects of electrospinning parameters on fiber diameter using XML techniques. The impact of parameters such as rotation speed, flow rate, voltage, and porosity on fiber diameter was examined in detail, and the complex relationships between these parameters were revealed. The results indicated that the RF model demonstrated the best performance. Evaluation on the test set yielded an RMSE of 129.3, MAPE of 0.371, R^2 of 0.54, and EVS of 0.59, suggesting that the RF model outperformed others in terms of both accuracy and interpretability. Hyperparameter optimization using BO improved the model performance. After optimization, the RMSE value decreased to 128.0, and the R^2 value increased to 0.555. The higher performance of BO can be attributed to its more efficient exploration of a broader hyperparameter space compared to GS, yielding optimal results with fewer trials. The study demonstrates that the effects of electrospinning parameters on fiber diameter are shaped by non-linear interactions. The RF model successfully captured these complex relationships, achieving high accuracy. Other regression models, particularly linear regression (Ridge and Lasso), and neighborhood-based methods (KNN), performed less effectively due to their inability to model complex relationships. Furthermore, SHAP analysis provided clearer insight into the effects of the parameters on fiber diameter. According to SHAP values, the rotation speed and flow rate had the strongest positive effects on fiber diameter. Increasing the rotation speed expanded the fiber diameter, while the flow rate also had an increasing effect. In contrast, the distance parameter had a positive but less pronounced effect on fiber diameter, reflected by lower SHAP values, although it remains significant. The voltage and temperature parameters negatively affected the fiber diameter. These findings contribute to the transparency of the model's decisions, allowing for a clearer understanding of how each parameter affects fiber diameter.

The limitations of the study include the small size of the dataset and the exclusive focus on studies using pure PCL. This may affect the model's generalizability. Future research with larger and more diverse datasets will yield more robust results. It is particularly important to focus on pure polymers in electrospinning processes. Although most of the literature emphasizes composite materials, studies on pure polymers are relatively scarce. The investigation of pure polymers is critical to gaining deeper insight and improving optimization methods. Future work investigating the effects of pure polymers on electrospinning processes will contribute significantly to advancing the knowledge in this area. In addition, a more comprehensive examination of the impact of solvents, environmental conditions, and other parameters on fiber morphology would be beneficial. Future studies should aim to enhance the optimization of electrospinning processes by focusing more on pure polymers.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability. Any additional information related to the study can be requested from the corresponding author.

References

[1] J. A Ilemobayo, O. Durodola, O. Alade, O. J Awotunde, A. T Olanrewaju, O. Falana, A. Ogungbire, A. Osinuga, D. Ogunbiyi, A. Ifeanyi, I. E Odezuligbo and O. E Edu, *Hyperparameter tuning in machine learning: a comprehensive review*, Journal of Engineering Research and Reports **26** (6), 388-395, 2024.

[2] A. Al-Abduljabbar and I. Farooq, *Electrospun polymer nanofibers: processing, properties, and applications*, Polymers **15** (1), 2023.

[3] N. Alharbi, A. Daraei, H. Lee and M. Guthold, *The effect of molecular weight and fiber diameter on the mechanical properties of single, electrospun pcl nanofibers*, Materials Today Communications **35**, 105773, 2023.

[4] E. Archer, M. Torretti and S. Madbouly, *Biodegradable polycaprolactone (pcl) based polymer and composites*, Physical Sciences Reviews **8** (11), 4391-4414, 2023.

[5] Z. Asvar, E. Mirzaei, N. Azarpira, B. Geramizadeh and M. Fadaie, *Evaluation of electrospinning parameters on the tensile strength and suture retention strength of polycaprolactone nanofibrous scaffolds through surface response methodology*, J Mech Behav Biomed Mater **75**, 369-378, 2017.

[6] M. Bartnikowski, T.R. Dargaville, S. Ivanovski and D.W. Hutmacher, *Degradation mechanisms of polycaprolactone in the context of chemistry, geometry and environment*, Progress in Polymer Science **96**, 1-20, 2019.

[7] D.M. Belete and M.D. Huchaiah, *Grid search in hyperparameter optimization of machine learning models for prediction of hiv/aids test results*, International Journal of Computers and Applications **44** (9), 875-886, 2022.

[8] O.O. Bifarin, *Interpretable machine learning with tree-based shapley additive explanations: application to metabolomics datasets for binary classification*, PLoS One **18** (5), e0284315, 2023.

[9] N.D. Bikiaris, I. Koumentakou, G. Michailidou, M. Kostoglou, M. Vlachou, P. Barmpalexis, E. Karavas and G.Z. Papageorgiou, *Investigation of molecular weight, polymer concentration and process parameters factors on the sustained release of the anti-multiple-sclerosis agent teriflunomide from poly(ε-caprolactone) electrospun nanofibrous matrices*, Pharmaceutics **14** (8), 2022.

[10] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A. Boulesteix, D. Deng and M. Lindauer, *Hyperparameter optimization: foundations, algorithms, best practices, and open challenges*, WIREs Data Mining and Knowledge Discovery **13** (2), e1484, 2023.

[11] J. Boelrijk, B. Pirok, B. Ensing and P. Forre, *Bayesian optimization of comprehensive two-dimensional liquid chromatography separations*, J Chromatogr A **1659**, 462628, 2021.

[12] A. Botchkarev, *A new typology design of performance metrics to measure errors in machine learning regression algorithms*, Interdisciplinary Journal of Information, Knowledge, and Management **14**, 045076, 2019.

[13] M. Chen, H. Michaud and S. Bhowmick, *Controlled vacuum seeding as a means of generating uniform cellular distribution in electrospun polycaprolactone (pcl) scaffolds*, J Biomech Eng **131** (7), 074521, 2009.

[14] Z. Chen, Z. Zhang, Y. Ouyang, Y. Chen, X. Yin, Y. Liu, H. Ying and W. Yang, *Electrospinning polycaprolactone/collagen fiber coatings for enhancing the corrosion resistance and biocompatibility of az31 mg alloys*, Colloids and Surfaces A Physicochemical and Engineering Aspects **662**, 2023.

[15] D. Chicco, M.J. Warrens and G. Jurman, *The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation*, PeerJ Computer Science **7**, e623, 2021.

[16] F. Croisier, A.S. Duwez, C. Jerome, A.F. Leonard, K.O. van der Werf, P.J. Dijkstra and M.L. Bennink, *Mechanical testing of electrospun pcl fibers*, Acta Biomater **8** (1), 218-24, 2012.

[17] J.R. Dias, A. Sousa, A. Augusto, P.J. Bartolo and P.L. Granja, *Electrospun polycaprolactone (pcl) degradation: an in vitro and in vivo study*, Polymers **14** (16), 2022.

[18] M.D. Edwards, G.R. Mitchell, S.D. Mohan and R.H. Olley, *Development of orientation during electrospinning of fibres of poly(ε-caprolactone)*, European Polymer Journal **46** (6), 1175-1183, 2010.

[19] S.J. Eichhorn and W.W. Sampson, *Statistical geometry of pores and statistics of porous nanofibrous assemblies*, Journal of the royal society Interface **2** (4), 309-318, 2005.

[20] B. Esteki, M. Masoomi, M. Moosazadeh and C. Yoo, *Data-driven prediction of janus/coreshell morphology in polymer particles: a machine-learning approach*, Langmuir **39** (14), 4943-4958, 2023.

[21] A.H. Etefagh and M.R. Razfar, *Bayesian optimization of 3d bioprinted polycaprolactone/magnesium oxide nanocomposite scaffold using a machine learning technique*, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture **238** (10), 1448-1462, 2024.

[22] Y. Fan, X. Miao, C. Hou, J. Wang, J. Lin and F. Bian, *High tensile performance of pla fiber-reinforced pcl composite via a synergistic process of strain and crystallization*, Polymer **270**, 125778, 2023.

[23] P. Francavilla, D.P. Ferreira, J.C. Araujo and R. Fangueiro, *Smart fibrous structures produced by electrospinning using the combined effect of pcl/graphene nanoplatelets*, Applied Sciences **11** (3), 2021.

[24] Z. Fu, D. Li, J. Cui, H. Xu, C. Yuan, P. Wang, B. Zhao and K. Lin, *Promoting bone regeneration via bioactive calcium silicate nanowires reinforced poly (ε-caprolactone) electrospun fibrous membranes*, Materials & Design **226**, 111671, 2023.

[25] K. Fujihara, M. Kotaki and S. Ramakrishna, *Guided bone regeneration membrane made of polycaprolactone/calcium carbonate composite nano-fibers*, Biomaterials **26** (19), 4139-47, 2005.

[26] S. Gautam, S.D. Purohit, H. Singh, A. Dinda, D.P. Potdar, C. Sharma, C. Chou and N. Mishra, *Surface modification of pcl-gelatin-chitosan electrospun scaffold by nano-hydroxyapatite for bone tissue engineering*, Materials Today Communications **34**, 105237, 2022.

[27] D. Gupta, J. Venugopal, M.P. Prabhakaran, V.R. Dev, S. Low, A.T. Choon and S. Ramakrishna, *Aligned and random nanofibrous substrate for the in vitro culture of schwann cells for neural tissue engineering*, Acta Biomater **5** (7), 2560-9, 2009.

[28] O. Ishii, M. Shin, T. Sueda and J.P. Vacanti, *In vitro tissue engineering of a cardiac graft using a degradable scaffold with an extracellular matrix-like topography*, The Journal of Thoracic and Cardiovascular Surgery **130** (5), 1358-1363, 2005.

[29] B. Joseph, A.J. John, J. Glamolija, D. Stojkovi, M. Sokovi, S. Lazovi, J. Kochupurackal, N. Kalarikkal and S. Thomas, *Processing and evaluation of the structure-properties of electrospun pcl/ zirconium nanoparticle scaffolds*, Materials Today Communications **34**, 104961, 2023.

[30] C. Kaliaperumal and A. Thulasisingh, *Electrospun polycaprolactone/chitosan/pectin composite nanofibre: a novel wound dressing scaffold*, Bulletin of Materials Science **46** (1), 23, 2023.

[31] L. Kalluri, M. Satpathy and Y. Duan, *Effect of electrospinning parameters on the fiber diameter and morphology of plga nanofibers*, Dent Oral Biol Craniofacial Res **4** (2), 2021.

[32] F. Kashani-Asadi-Jafari, A. Parhizgar and A. Hadjizadeh, *Magnetic-field-assisted emulsion electrospinning system: designing, assembly, and testing for the production of pcl/gelatin coreshell nanofibers*, Fibers and Polymers **24**, 515-523, 2023.

[33] K.A.G. Katsogiannis, G.T. Vladisavljevi and S. Georgiadou, *Porous electrospun polycaprolactone fibers: effect of process parameters*, Journal of Polymer Science Part B: Polymer Physics **54** (18), 1878-1888, 2016.

[34] J.W. Kim, S. Park, K. Park and B. Kim, *Non-toxic natural additives to improve the electrical conductivity and viscosity of polycaprolactone for melt electrospinning*, Applied Sciences **13** (3), 2023.

[35] J. Ko, S. Jun, J. Lee, P. Lee and M. Jun, *Effects of molecular weight and temperature on fiber diameter of poly(ε-caprolactone) melt electrospun fiber*, Journal of the Korean Society of Manufacturing Technology Engineers **24**, 160-165, 2015.

[36] J. Lévesque, C. Gagné and R. Sabourin, *Bayesian hyperparameter optimization for ensemble learning*, Proceedings Of The Thirty-second Conference On Uncertainty In Artificial Intelligence, 437446, Arlington, Virginia, USA, 2016.

[37] W.J. Li, K.G. Danielson, P.G. Alexander and R.S. Tuan, *Biological response of chondrocytes cultured in three-dimensional nanofibrous poly(ε-caprolactone) scaffolds*, J Biomed Mater Res A **67** (4), 1105-14, 2003.

[38] M. Li, H. Sun, Y. Huang and H. Chen, *Shapley value: from cooperative game to explainable artificial intelligence*, Autonomous Intelligent Systems **4** (1), 2, 2024.

[39] W.J. Li, R. Tuli, C. Okafor, A. Derfoul, K.G. Danielson, D.J. Hall and R.S. Tuan, *A three-dimensional nanofibrous scaffold for cartilage tissue engineering using human mesenchymal stem cells*, Biomaterials **26** (6), 599-609, 2005.

[40] W.J. Li, R. Tuli, X. Huang, P. Laquerriere and R.S. Tuan, *Multilineage differentiation of human mesenchymal stem cells in a three-dimensional nanofibrous scaffold*, Biomaterials **26** (25), 5158-66, 2005.

[41] Q. Liang, A.E. Gongora, Z. Ren, A. Tiihonen, Z. Liu, S. Sun, J.R. Deneault, D. Bash, F. Mekki-Berrada, S.A. Khan, K. Hippalgaonkar, B. Maruyama, K.A. Brown, J. Fisher Iii and T. Buonassisi, *Benchmarking the performance of bayesian optimization across multiple experimental materials science domains*, npj Computational Materials **7** (1), 188, 2021.

[42] F.J. Lopez-Flores, J.A. Ornelas-Guillen, A. Perez-Nava, J.B. Gonzalez-Campos and J.M. Ponce-Ortega, *Data-driven machine learning approach for modeling the production and predicting the characteristics of aligned electrospun nanofibers*, Industrial & Engineering Chemistry Research **63** (22), 9904-9913, 2024.

[43] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.I. Lee, *From local explanations to global understanding with explainable ai for trees*, Nat Mach Intell **2** (1), 56-67, 2020.

[44] M. Ma, H. Zhou, S. Gao, N. Li, W. Guo and Z. Dai, *Analysis and prediction of electrospun nanofiber diameter based on artificial neural network*, Polymers **15** (13), 2023.

[45] E. Malikmammadov, T.E. Tanir, A. Kiziltay, V. Hasirci and N. Hasirci, *Pcl and pcl-based materials in biomedical applications*, J Biomater Sci Polym Ed **29** (7-9), 863-893, 2018.

[46] T.B. Martin and D.J. Audus, *Emerging trends in machine learning: a polymer perspective*, ACS Polymers Au **3** (3), 239-258, 2023.

[47] A. Mishra, V.S. Jatti, E.M. Sefene and S. Paliwal, *Explainable artificial intelligence (xai) and supervised machine learning-based algorithms for prediction of surface roughness of additively manufactured polylactic acid (pla) specimens*, Applied Mechanics **4** (2), 668-698, 2023.

[48] R. Mitchell, E. Frank and G. Holmes, *Gputreeshap: massively parallel exact calculation of shap scores for tree ensembles*, PeerJ Comput Sci **8**, e880, 2022.

[49] M. Nemeth, D. Borkin and G. Michalconok, *The Comparison Of Machine-learning Methods Xgboost And Lightgbm To Predict Energy Development*, 208-215, 2019.

[50] R.M. Nezarati, M.B. Eifert and E. Cosgriff-Hernandez, *Effects of humidity and solution viscosity on electrospun fiber morphology*, Tissue Eng Part C Methods **19** (10), 810-9, 2013.

[51] D.R. Nisbet, A.E. Rodda, M.K. Horne, J.S. Forsythe and D.I. Finkelstein, *Neurite infiltration and cellular response to electrospun polycaprolactone scaffolds implanted into the brain*, Biomaterials **30** (27), 4573-80, 2009.

[52] C. Nnaji and U. Nwodo, *Predicting customer churn in the telecommunication industry using machine learning algorithms: performance comparison with logistic regression,random forest, and gradient boosting techniques.*, Machine Learning **22**, 3-66, 2022.

[53] N. Obregon, V. Agubra, M. Pokhrel, H. Campos, D. Flores, D. De la Garza, Y. Mao, J. Macossay and M. Alcoutlabi, *Effect of polymer concentration, rotational speed, and solvent mixture on fiber formation using forcespinningő*, Fibers **4** (2), 2016.

[54] E. Pektok, B. Nottelet, J. Tille, R. Gurny, A. Kalangos, M. Moeller and B.H. Walpoth, *Degradation and healing characteristics of small-diameter poly(ε-caprolactone) vascular grafts in the rat systemic arterial circulation*, Circulation **118** (24), 2563-2570, 2008.

[55] M.N. Pervez, W.S. Yeo, M.M.R. Mishu, M.E. Talukder, H. Roy, M.S. Islam, Y. Zhao, Y. Cai, G.K. Stylios and V. Naddeo, *Electrospun nanofiber membrane diameter prediction using a combined response surface methodology and machine learning approach*, Scientific Reports **13** (1), 9679, 2023.

[56] Q.P. Pham, U. Sharma and A.G. Mikos, *Electrospun poly(ε-caprolactone) microfiber and multilayer nanofiber/microfiber scaffolds: characterization of scaffolds and measurement of cellular infiltration*, Biomacromolecules **7** (10), 2796-805, 2006.

[57] M.P. Prabhakaran, J. Venugopal, C.K. Chan and S. Ramakrishna, *Surface modified electrospun nanofibrous scaffolds for nerve tissue engineering*, Nanotechnology **19** (45), 455102, 2008.

[58] M. R and R. Rohini, *Lasso: a feature selection technique in predictive modeling for machine learning*, 18-20, 2016.

[59] O. Rainio, J. Teuho and R. Klen, *Evaluation metrics and statistical tests for machine learning*, Scientific Reports **14** (1), 6086, 2024.

[60] S. Ramazani and M. Karimi, *Investigating the influence of temperature on electrospinning of polycaprolactone solutions*, e-Polymers **14** (5), 323-333, 2014.

[61] R. Rodriguez-Perez and J. Bajorath, *Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions*, J Comput Aided Mol Des **34** (10), 1013-1026, 2020.

[62] K. Safjan, *Explaining ai - the key differences between lime and shap methods*, Krystian's Safjan Blog, 2023.

[63] V. Salaris, A. Leones, D. Lopez, J.M. Kenny and L. Peponi, *A comparative study on the addition of mgo and mg(oh)2 nanoparticles into pcl electrospun fibers*, Macromolecular Chemistry and Physics **224** (1), 2200215, 2023.

[64] A.M. Salih, Z. Raisi-Estabragh, I.B. Galazzo, P. Radeva, S.E. Petersen, K. Lekadir and G. Menegaz, *A perspective on explainable artificial intelligence methods: shap and lime*, Advanced Intelligent Systems **7** (1), 2400304, 2025.

[65] S. Sarma, A.K. Verma, S.S. Phadkule and M. Saharia, *Towards an interpretable machine learning model for electrospun polyvinylidene fluoride (pvdf) fiber properties*, Computational Materials Science **213**, 111661, 2022.

[66] R. Scaffaro, M. Gammino and A. Maio, *Wet electrospinning-aided self-assembly of multifunctional go-cnt@pcl core-shell nanocomposites with spider leg bioinspired hierarchical architectures*, Composites Science and Technology **221**, 109363, 2022.

[67] R. Schofield, B. Maciejewska, S. Dong, G. Tebbutt, D. McGurty, R. Bonilla, H. Assender and N. Grobert, *Driving fiber diameters to the limit: nanoparticle-induced diameter reductions in electrospun photoactive composite nanofibers for organic photovoltaics*, Advanced Composites and Hybrid Materials **6**, 2023.

[68] M. Shie Karizmeh, S.A. Poursamar, A. Kefayat, Z. Farahbakhsh and M. Rafienia, *An in vitro and in vivo study of pcl/chitosan electrospun mat on polyurethane/propolis foam as a bilayer wound dressing*, Mater Sci Eng C Mater Biol Appl **135**, 112667, 2022.

[69] G. Simsek Gunduz, *Investigation of the effect of needle diameter and the solution flow rate on fiber morphology in the electrospinning method*, Fibres & Textiles in Eastern Europe **31**, 2023.

[70] M. Sivan, D. Madheswaran, J. Valtera, E.K. Kostakova and D. Lukas, *Alternating current electrospinning: the impacts of various high-voltage signal shapes and frequencies on the spinnability and productivity of polycaprolactone nanofibers*, Materials & Design **213**, 110308, 2022.

[71] T. Subbiah, G.S. Bhat, R.W. Tock, S. Parameswaran and S.S. Ramkumar, *Electrospinning of nanofibers*, Journal of Applied Polymer Science **96** (2), 557-569, 2005.

[72] B. Subeshan, A. Atayo and E. Asmatulu, *Machine learning applications for electrospun nanofibers: a review*, Journal of Materials Science **59** (31), 14095-14140, 2024.

[73] S. Sukpancharoen, T. Wijakmatee, T. Katongtung, K. Ponhan, N. Rattanachoung and S. Khojitmate, *Data-driven prediction of electrospun nanofiber diameter using machine learning: a comprehensive study and web-based tool development*, Results in Engineering **24**, 102826, 2024.

[74] R.S. Tigli, N.M. Kazaroglu, B. Mavis and M. Gumusderelioglu, *Cellular behavior on epidermal growth factor (egf)-immobilized pcl/gelatin nanofibrous scaffolds*, J Biomater Sci Polym Ed **22** (1-3), 207-23, 2011.

[75] A.D. Tsareva, V.S. Shtol, D.V. Klinov and D.A. Ivanov, *Electrospinning for biomedical applications: an overview of material fabrication techniques*, Surfaces **8** (1), 2025.

[76] J. Ukwaththa, S. Herath and D. Meddage, *A review of machine learning (ml) and explainable artificial intelligence (xai) methods in additive manufacturing (3d printing)*, Materials Today Communications **41**, 110294, 2024.

[77] J. Venugopal, L.L. Ma, T. Yong and S. Ramakrishna, *In vitro study of smooth muscle cells on polycaprolactone and collagen nanofibrous matrices*, Cell Biol Int **29** (10), 861-7, 2005.

[78] J. Venugopal, Y.Z. Zhang and S. Ramakrishna, *Fabrication of modified and functionalized polycaprolactone nanofibre scaffolds for vascular tissue engineering*, Nanotechnology **16** (10), 2138-42, 2005.

[79] S. Vinay, *Standardization in machine learning*, 2021.

[80] E. Vivas, H. Allende-Cid and R. Salas, *A systematic review of statistical and machine learning methods for electrical power forecasting with reported mape score*, Entropy **22** (12), 1412, 2020.
[81] C. Wan and B. Chen, *Poly(ε-caprolactone)/graphene oxide biocomposites: mechanical properties and bioactivity*, Biomed Mater **6** (5), 055010, 2011.
[82] C. Wongoutong, *The impact of neglecting feature scaling in k-means clustering*, PLoS One **19** (12), e0310839, 2024.
[83] Y. Xie, Q. Fang, H. Zhao, Y. Li, Z. Lin and J. Chen, *Effects of six processing parameters on the size of pcl fibers prepared by melt electrospinning writing*, Micromachines **14** (7), 2023.
[84] S. Xie and R.T. Ogden, *Functional support vector machine*, Biostatistics **25** (4), 1178-1194, 2024.
[85] S. Yadav and S. Shukla, *Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification*, 2016 Ieee 6th International Conference On Advanced Computing (iacc), 78-83, 2016.
[86] J. Yang, Fast treeshap: accelerating shap value computation for trees, 2022.
[87] H. Yildirim and M.. Ozkale, *The performance of elm based ridge regression via the regularization parameters*, Expert Systems with Applications **134**, 225-233, 2019.
[88] H. Yoshimoto, Y.M. Shin, H. Terai and J.P. Vacanti, *A biodegradable nanofiber scaffold by electrospinning and its potential for bone tissue engineering*, Biomaterials **24** (12), 2077-82, 2003.
[89] H. Yu, J. Jang, T. Kim, H. Lee and H. Kim, *Apatite-mineralized polycaprolactone nanofibrous web as a bone tissue regeneration substrate*, Journal of Biomedical Materials Research Part A **88A** (3), 747-754, 2009.
[90] Z. Zhang, *Introduction to machine learning: k-nearest neighbors*, Ann Transl Med **4** (11), 218, 2016.
[91] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society Series B: Statistical Methodology **67** (2), 301-320, 2005.

Appendix A. Final Hyperparameter Values Under 3-fold CV

Table 10. Representative Final Hyperparameter Values for All Models Under 3-fold CV.

Model	GS	BO
XGBoost	n_estimators=180 learning_rate=0.01 max_depth=6 subsample=0.8	n_estimators=200 learning_rate=0.009 max_depth=8 subsample=0.85
RF	n_estimators=180 max_depth=10 min_samples_split=2	n_estimators=210 max_depth=12 min_samples_split=2
GB	n_estimators=120 learning_rate=0.01 max_depth=5	n_estimators=150 learning_rate=0.008 max_depth=6
DecisionTree	max_depth=8 min_samples_split=2	max_depth=10 min_samples_split=2
KNN	n_neighbors=5 leaf_size=30	n_neighbors=7 leaf_size=25
LightGBM	n_estimators=120 num_leaves=31 learning_rate=0.01	n_estimators=150 num_leaves=40 learning_rate=0.009
SVR	C=80 gamma=0.01 kernel=rbf	C=100 gamma=0.008 kernel=rbf
ElasticNet	alpha=0.12 l1_ratio=0.5	alpha=0.10 l1_ratio=0.55
Ridge	alpha=1.2	alpha=1.0
Lasso	alpha=0.006	alpha=0.005

Appendix B. Final Hyperparameter Values Under 10-fold CV

Table 11. Representative Final Hyperparameter Values for All Models Under 10-fold CV

Model	GS	BO
XGBoost	n_estimators=220 learning_rate=0.01 max_depth=7 subsample=0.8	n_estimators=250 learning_rate=0.008 max_depth=9 subsample=0.85
RF	n_estimators=220 max_depth=10 min_samples_split=2	n_estimators=250 max_depth=12 min_samples_split=2
GB	n_estimators=160 learning_rate=0.01 max_depth=5	n_estimators=200 learning_rate=0.008 max_depth=6
DecisionTree	max_depth=10 min_samples_split=2	max_depth=12 min_samples_split=2
KNN	n_neighbors=6 leaf_size=30	n_neighbors=8 leaf_size=25
LightGBM	n_estimators=160 num_leaves=31 learning_rate=0.01	n_estimators=190 num_leaves=45 learning_rate=0.008
SVR	C=120 gamma=0.01 kernel=rbf	C=140 gamma=0.008 kernel=rbf
ElasticNet	alpha=0.11 l1_ratio=0.50	alpha=0.09 l1_ratio=0.55
Ridge	alpha=1.0	alpha=0.85
Lasso	alpha=0.005	alpha=0.004