ML Approaches for Predicting Mechanical Properties of Artificial Spider Silk from Spinning Conditions

1st Erik Lidbjörk
Division of Robotics Perception and Learning
KTH Royal Institute of Technology
Stockholm, Sweden
0009-0006-6794-7030, eriklidb@kth.se

2nd Hedvig Kjellström
Division of Robotics Perception and Learning
KTH Royal Institute of Technology
Stockholm, Sweden
0000-0002-5750-9655, hedvig@kth.se

3rd Neeru Dubey
Division of Robotics Perception and Learning
KTH Royal Institute of Technology
Stockholm, Sweden
needub@kth.se

***Abstract*—Sustainable alternatives to petroleum-based fibers are urgently needed. Spider silk offers exceptional strength, extensibility, toughness, and biodegradability, but large-scale natural production is unfeasible. Recombinant spidroins can be processed into fibers, yet their mechanical performance often lags behind native silk. Fiber properties depend strongly on spinning conditions such as protein sequence, concentration, solvent, and method. Here, we analyze a curated dataset linking processing parameters to fiber mechanics. Machine learning models are applied to predict five mechanical properties and assess the contribution of sequence information. The results provide actionable insights for optimizing artificial spider silk toward high-performance, sustainable applications.**

***Index Terms*—Keywords, ...**

I. Introduction

The growing demand for sustainable and high-performance materials has intensified the search for alternatives to petroleum-derived synthetic fibers. Conventional fibers, such as nylon and polyester, dominate the textile and materials industry; yet their production relies on fossil resources and contributes significantly to environmental pollution [1]. A promising bio-based alternative is spider silk, a proteinaceous fiber renowned for its exceptional combination of strength, extensibility, toughness, and biodegradability [2], [3].

Natural spider silk, however, is not a scalable resource. Spiders produce only small amounts of silk and exhibit territorial and cannibalistic behavior, making large-scale farming impractical [4]. To address this limitation, research has focused on producing *artificial spider silk* by recombinantly expressing spider silk proteins (spidroins) in heterologous hosts such as bacteria, yeast, or mammalian cells [5], [6]. Once purified, spidroins can be solubilized and spun into fibers through laboratory techniques designed to mimic spiders' natural spinning process [7].

The fiber properties of recombinant silk are strongly influenced by the *spinning conditions*, a term that encompasses parameters such as protein sequence, concentration, solvent composition, extrusion method, and post-spinning treatments [8], [9]. While substantial progress has been made, artificial silk fibers generally do not yet match the mechanical performance of natural spider silk [10]. A deeper understanding of how spinning conditions shape fiber properties is, therefore, critical for optimizing production and tailoring silk to desired applications.

In this project, we analyze experimental data generated by Rising's group at the Swedish University of Agricultural Sciences (SLU), which systematically records both spinning conditions and mechanical outcomes of recombinant spider silk fibers [11]. The experimental setup used by the Rising group (1) used to produce the dataset, involves extruding a concentrated spidroin solution through a glass capillary into an aqueous spinning bath and collecting the resulting fibers on a rotating wheel. We propose to apply machine learning models to this dataset with two objectives: (i) to predict mechanical properties—tensile strength, strain, toughness, diameter, and Young's modulus—based on spinning parameters, and (ii) to assess the role of spidroin sequence information in improving model performance. By uncovering correlations between production conditions and fiber mechanics, this work aims to provide insights that support the development of bio-based fibers with properties approaching those of natural spider silk.

Major contributions of this work are as follows:

- Identification of the most influential spinning conditions for each of the five key mechanical properties (tensile strength, strain, toughness, diameter, and Young's modulus).
- Exploration of the relationship between amino acid sequence patterns and the resulting mechanical properties of spun silk fibers.

The remainder of the paper is organized as follows: Section II reviews the related literature and existing approaches. Section III details the proposed framework and experimental

Identify applicable funding agency here. If none, delete this.

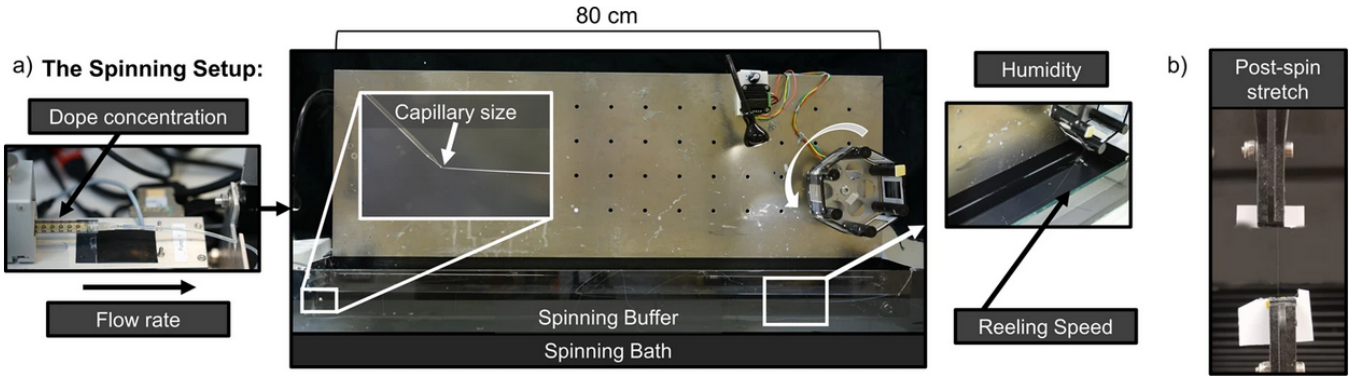

Fig. 1. (a) A concentrated aqueous dope of recombinant spidroins is extruded through a glass capillary into an aqueous spinning bath and collected on a rotating wheel. Key adjustable parameters include dope concentration, flow rate, capillary size, bath composition and pH, reeling speed, and ambient humidity. (b) Post-spin stretching is performed using a tensile tester to evaluate mechanical reinforcement [12].

setup. Section IV presents the findings and discusses their implications. Finally, Section V concludes the paper and outlines directions for future work.

II. Related Works

Engineering artificial spider silk has advanced along two complementary fronts: sequence-driven modeling and experimental control of processing. On the computational side, recent work has framed spidroin design as a conditional generative modeling problem. Lu *et al.* [13] developed *SilkomeGPT*, a transformer pre-trained on large protein corpora and fine-tuned on $\sim$1,000 MaSp sequences with paired fiber-level labels, enabling both forward (sequence$\rightarrow$properties) and inverse (target$\rightarrow$sequence) design. In parallel, Dubey *et al.* [14] distilled ProtGPT2 and fine-tuned it on curated MaSp repeats (including 572 with experimental labels), demonstrating property-conditioned generation and sequence-to-property prediction. These efforts build on protein language models showing transferable embeddings for structure/function learning, notably ProtTrans [15] and ESM, which achieved state-of-the-art structure and function inference when scaled to hundreds of millions of sequences [16].

Experimental studies have established that processing parameters exert dominant control over fiber mechanics. Schmuck *et al.* [12] tested $>$1,000 biomimetically spun fibers across 92 conditions, finding reeling speed, bath chemistry, and post-spin stretching to be primary determinants of tensile variability. Natural spinning and forcible-silking studies corroborate these dependencies: Young *et al.* [17] related spinning rate to molecular alignment using Raman spectroscopy, while Pérez-Rigueiro *et al.* [18] showed that silking force correlates tightly with stiffness and extensibility. Wet-spinning experiments further demonstrated that solvent composition and post-draw conditions can be tuned to yield highly elastic fibers [19]. Methodological differences (strain rate, humidity, gauge normalization) can also bias reported values, highlighting the need for standardization [20]. Molecular simulations now bridge sequence-level design and atomistic mechanics [21].

Convergence is emerging between sequence engineering, scalable production, and standardized process control. Comprehensive reviews synthesize strategies for high-performance artificial spider silk; covering protein architecture, spinning dope formulation, post; draw protocols, and composite reinforcement; while calling for integrated datasets and unified evaluation [22]. Recent advances include engineered mini-spidroins reaching dragline-like toughness under biomimetic spinning [23], high-yield recombinant production of super-soluble variants [24], and artificial spinning setups that reproduce native gland triggers such as pH gradient, shear, and dehydration [7], [25], [26]. Collectively, these works underscore the opportunity for predictive models that fuse frozen PLM embeddings with detailed spinning and process features; an approach our study directly explores.

III. Methodology

The tabular dataset produced by Rising's group [11] originally contained 23 columns (features) for the spinning conditions, as well as 734 rows (number of sets of spinning conditions). The features contained a mix of numerical and categorical data, with a high frequency of missing data in certain columns ($\sim$66% for pump pressure, $\sim$3 for spinning temperature and continuous spinning, $\sim$1% for spinning humidity, and $<$1% for the remaining columns). For each spinning condition (row), Rising's group took around ten measurements of the five mechanical properties: tensile strength, strain, toughness, diameter, and Young's modulus, for each experiment. To maximize the number of data points, we considered each spinning condition—mechanical property target as a single input-output example, instead of computing a statistic for the batch (e.g., mean). Due to time constraints, the Rising group only labeled a fraction of the dataset, which limited the amount of data for training and evaluation. For some data points, the mean for each measurement of the mechanical properties was recorded. This resulted in a total of 2190 data points. The dataset was then preprocessed by removing certain columns, such as dates. The data was then preprocessed by cleaning, standardizing, unit conversion, and

converting certain numerical data into categorical data. The final dataset included the following parameters: The dataset contains the following features: the type of protein spun to create the silk, concentration ($\mathrm{mg/mL}$), type of spinning device, type of extrusion device, number of baths used (one or two), spinning bath temperature ($^{\circ}\mathrm{C}$), type of spinning buffer, spinning bath pH, spinning bath concentration $\mathrm{mmol/L}$, NaCl $\mathrm{mmol/L}$ concentration, type of capillary device, The protein sequences for each type of protein in the dataset were obtained, processed, and converted into fixed-length vector embeddings using ESM-2 (esm2_t30_150M_UR50D), a large transformer-based protein language model (PLM) pre-trained on millions of protein sequences [27], which were then reduced from 640 dimensions to 15 with PCA. The data for the mechanical properties were mapped onto $[0, 1]$ using min-max (linear) scaling, thus making them unitless, and thus making the model predictions themselves, as well as resulting metrics, unitless. We will present the statistics using the scaled data only for easier comparisons between the mechanical properties.

We trained predictive models using gradient boosting tree estimators [28]. We opted for gradient boosting instead of deep neural networks due to their negligible performance difference [29] on tabular data, along with the relatively low sample size, which excluded deep learning models. We choose the HistGradientBoostingRegressor (HGBR) implementation from scikit-learn [30], selected for its efficiency on tabular data, support for heterogeneous features, and ability to handle missing values natively. One HGBR was trained per mechanical property. In total, the model consists of five HGBR estimators, one for each mechanical property. Model performance was evaluated using nested cross-validation, with three outer folds and five inner folds, with the inner loop dedicated to hyperparameter optimization via Optuna [31] and the outer loop providing unbiased estimates of predictive accuracy. We used grouped k-fold cross-validation grouped by the sample batches to avoid mixing of mechanical property values produced from the same experiment to appear in different folds, since we assume these to be dependent. Tuned parameters for each HGBR were comprised of: learning rate, the maximum number of leaves for each tree, the minimum number of samples per leaf, and L_2 regularization. The ranges sampled from for each hyperparameters, as well as a presentation of the values for the final models are presented in Table I, in the supplementary material. The loss function, which was minimized during training, was the squared error. The final trained model is then composed of a set of three HGBRs, one for each outer cross-validation fold. One set was trained on the pre-processed spinning condition data, and another set on the same data, with the PLM-derived vector embeddings appended to each datapoint, matching its corresponding protein. We will refer to these models as A and B, respectively, throughout the rest of the paper.

Evaluation metrics comprised root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and Pearson correlation coefficient (PCC). The impact of the PLM-derived embeddings on the model was analyzed using a paired sample Student's t-test between (scaled) mechanical property predictions between the two models (A and B). To interpret model predictions and assess the relative influence of input variables, we employed SHAP (SHapley Additive exPlanations) [32] for each mechanical property individually. Lastly, we performed an ablation study, where we trained an additional model only on the top-$K \in \{1, 2, ..., 10\}$ features, based on the SHAP-value ranking averaged across all mechanical properties.

IV. Results

Table I reports the performance of the model trained on the spinning condition data (A), and the model trained on the spinning conditions data and PLM-derived vector embeddings (B), and their difference ($B - A$) across the four evaluation metrics: root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and Pearson correlation coefficient (PCC). The metrics are calculated from the out-of-sample predictions (aggregated from all outer folds) from both models. Results are presented for each of the five mechanical properties. The best performing estimators for both models were for the diameter property, which is evident by a low RMSE and MAE, and a higher R^2 and PCC. Although the Young's modulus models achieved relatively low RMSE and MAE, the R^2 score was near zero, and the PCC was low. This suggests that the estimator predominantly predicts values close to the mean, achieving low absolute error because of the limited variance, but failing to explain variation across samples. This likely reflects either a weak predictive signal in the features or substantial label noise. The differences between the metrics for both models are generally quite small (around two or more orders of magnitude); however, we can see that the starkest differences between models are for the yield strength, where model B performs better, as evident by a lower RMSE and MAE, and higher R^2 and PCC.

To assess how well the models capture the distribution of target values, Fig 2. shows violin plots for the ground truth, Model A predictions, and Model B predictions for each property. The figure allows visual comparison of whether the predicted distributions align with the empirical distribution of the data, and whether systematic biases (e.g., underestimation of extremes, overly sharp peaks) are present. We can see that the target distributions for models A and B are very similar. The overall shape between ground-truth and prediction distributions looks quite similar, but they differ in the details. Most notably, the ground truth distributions are considerably wider, and also have present peaks at the low and, particularly, high ends. This is likely due to the noise in the target labels. As mentioned in chapter III, most datapoints came in batches where multiple fibers were measured coming from the same experiment, which were all considered a single data point due to data scarcity. The HGBR estimators, which are deterministic, will always predict the same output given the same input. The model outputs will likely ignore outliers and instead predict closer to the mean value of the batches. Thus, due to the methodology, the ground-truth distribution

TABLE I

COMPARISON OF THE METRICS ROOT MEAN SQUARED ERROR (RMSE), MEAN ABSOLUTE ERROR (MAE), COEFFICIENT OF DETERMINATION (R^2), AND PEARSON CORRELATION COEFFICIENT (PCC) FOR EACH MECHANICAL PROPERTY, BETWEEN THE SET OF MODELS TRAINED WITH (B), WITHOUT (A) PLM-DERIVED VECTOR EMBEDDINGS, AND DIFFERENCE $(B - A)$.

	RMSE			MAE			R^2			PCC		
	A	B	$B-A$	A	B	$B-A$	A	B	$B-A$	A	B	$B-A$
Strength	9.93E-02	9.69E-02	-2.43E-03	7.24E-02	6.99E-02	-2.43E-03	2.41E-01	2.78E-01	3.67E-02	5.18E-01	5.31E-01	1.35E-02
Strain	2.15E-01	2.15E-01	-7.52E-04	1.60E-01	1.60E-01	-6.42E-04	3.21E-01	3.25E-01	4.74E-03	5.78E-01	5.81E-01	3.32E-03
Toughness	1.55E-01	1.56E-01	1.08E-03	1.20E-01	1.20E-01	6.77E-04	2.00E-01	1.89E-01	-1.11E-02	4.58E-01	4.39E-01	-1.86E-02
Diameter	7.51E-02	7.50E-02	-7.97E-05	5.02E-02	5.00E-02	-1.84E-04	5.25E-01	5.26E-01	1.01E-03	7.27E-01	7.29E-01	2.12E-03
Young's modulus	7.26E-02	7.26E-02	2.06E-05	5.40E-02	5.34E-02	-5.75E-04	9.66E-03	9.10E-03	-5.63E-04	2.21E-01	2.27E-01	5.32E-03

is expected to have considerably greater variance than the target distributions. Table II shows the mean, sample standard deviation (SD), and coefficient of variation (mean / SD) (CV) for each property. We observe high CVs over 39% for all properties, meaning that property measurements are highly variable relative to their means, which indicates a high degree of noise in the labels. Fig 1, in the supplementary, presents the mean, SD, and CV, per experiment batch. Additionally, scatter diagrams of ground-truth vs. predictions (Fig 4) per target, from both models, as well as scatter diagrams of residuals (Fig 5), are also present in the supplementary material, which provides additional context on model prediction and signs of noise in the data. By visual inspection, some of the results in Table I mentioned can be corroborated by Fig 2. Namely, we can see that the distributions for the diameter, which was the best estimated property, look quite similar. On the other hand, we see that prediction distributions for Young's modulus look very narrow, which indicates low variance for the prediction, which makes sense considering the low R^2 and PCC.

TABLE II

SAMPLE MEAN, SAMPLE STANDARD DEVIATION (SD), AND COEFFICIENT OF VARIATION (CV) (SD/MEAN), FOR EACH PROPERTY OF THE MINMAX SCALED DATA. A CV $>39\%$ IMPLIES HIGH VARIABILITY IN THE DATA.

	Strength	Strain	Toughness	Young's modulus	Diameter
Mean	2.58E-01	3.96E-01	2.52E-01	1.91E-01	1.86E-01
SD	1.14E-01	2.61E-01	1.73E-01	1.09E-01	7.29E-02
CV	4.41E-01	6.60E-01	6.89E-01	5.69E-01	3.92E-01

Feature contributions were assessed using SHAP values. Figure 3 presents beeswarm plots of SHAP values from model B, for the top five ranked features, and the total sum of the remaining features, for each property, showing both importance ranking and directionality of feature effects. The sub-figure on the bottom right summarizes the averaged absolute SHAP importance across all properties, highlighting the top five most influential features overall, as well as the sum of the remaining features. Figure 2 in the supplementary presents the corresponding Figure for model A. The absolute magnitude reflects the strength of the contribution, with larger values corresponding to greater influence on the prediction, which is how the feature importance ranking is decided. The SHAP-values were obtained using out-of-sample data for model B; a figure showcasing SHAP feature importance ranking for model A is presented in Figure X, in the supplementary material. This analysis provides insight into which experimental or sequence-related variables have the strongest impact on predicting mechanical properties. Positive SHAP values indicate feature contributions that increase the predicted property, while negative values decrease it [32]. Colors represent feature magnitude (red = high, blue = low). For categorical variables (e.g., protein type), the color coding does not reflect a meaningful scale due to ordinal encoding. The top 5 ranking averaged between properties presented NaCl amount of substance, spinning temperature, protein type, spinning bath pH, and reeling speed. These five features contribute to $\sim$56% (0.09/0.16 SHAP value) of the total SHAP impact. We can see that the amount of NaCl has the highest SHAP value on average, having a high ranking for both toughness and strain. Spinning bath pH is also high ranked, having a top-five ranking for each property except diameter. The protein type is also deemed important, appearing in the top five features for each property (but never being number one).

To assess whether the predictions from Models A and B differ significantly, we applied paired Student's t-tests on the out-of-sample results for each property. Table III reports the test statistics, p-values, and 95% confidence intervals for the mean differences $(A - B)$. The hypotheses are formulated as:

$$H_0 : \mu_{A-B} = 0 \qquad \text{(no mean difference)} \tag{1}$$

$$H_1 : \mu_{A-B} \neq 0 \qquad \text{(non-zero mean difference)} \tag{2}$$

A paired t-test is appropriate here because the distributions of the paired differences (shown in Fig 3 in the supplementary material) are approximately normal and symmetric. Using a significance level of 5%, we find statistically significant differences for all properties except strain. This indicates that incorporating frozen embeddings has a measurable, though small, impact on predictions. However, the effect sizes are consistently small (Cohen's $d < 0.2$), suggesting that the observed differences, while statistically significant, are of limited practical importance. Table I further indicates that Model B tends to outperform Model A for strength and diameter.

To evaluate the robustness of the models to feature reduction, we conducted an ablation study by training a model on the top-$K \in \{1, 2, ..., K\}$ SHAP-ranked features averaged across all properties. The reduced-feature models were retrained using the same outer-fold hyperparameters as model B. Fig 4 depicts RMSE vs. K. We can see that for most

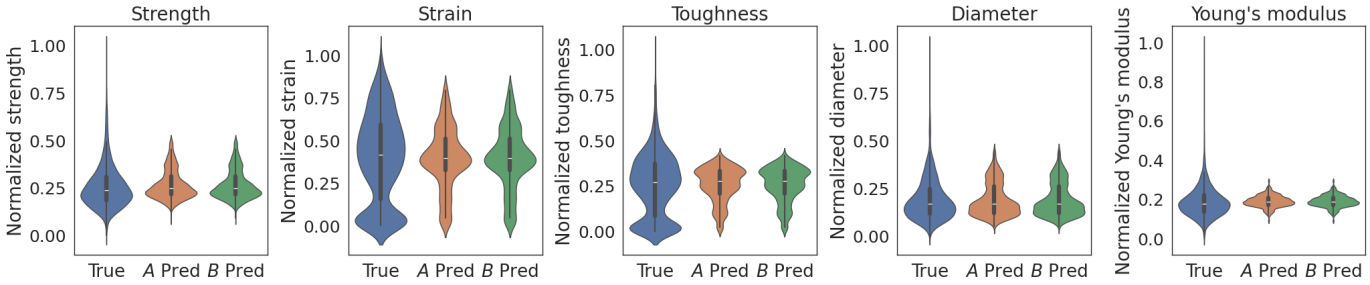

Fig. 2. The ground-truth distribution of data for each mechanical property vs. the predicted distributions of mechanical properties yielded by the model trained on the spinning conditions data only (A) and the model trained on spinning conditions data and PLM-derived vector embeddings (B), on unseen data. The distributions are presented as violin plots (kernel density estimations). In general, the actual distribution is wider than the predicted, likely due to label noise. Noisy outliers are also evident by the small peaks at the edges of the data distributions.

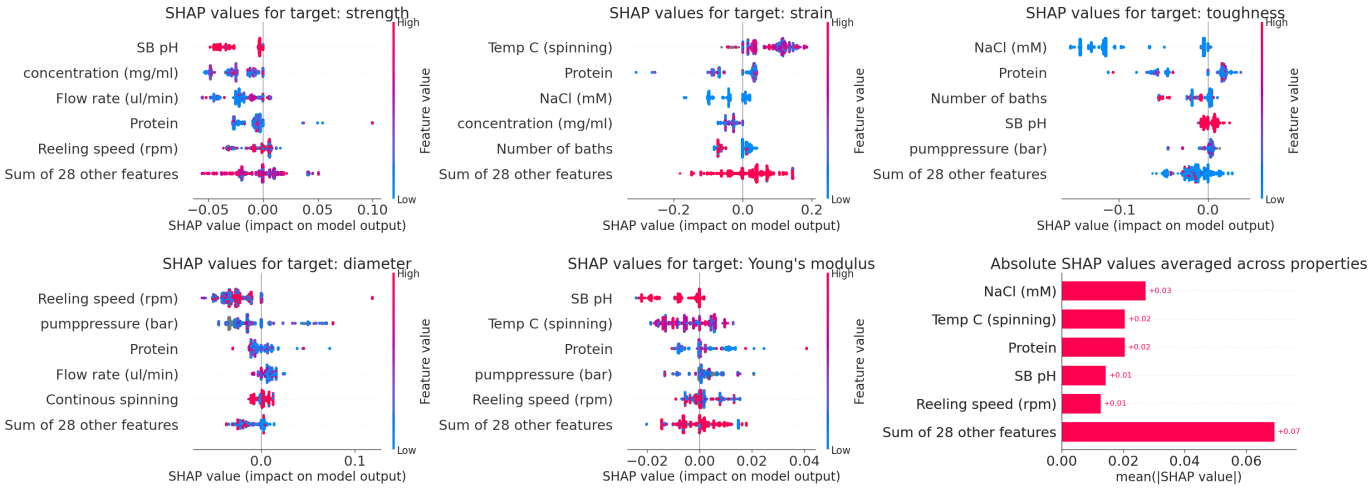

Fig. 3. Top five feature importance ranking for each mechanical property, derived from model B. Ranking is decided based on the total sum of absolute value SHAP-values for each feature. A high/low SHAP-value (position on the x-axis) indicates positive/negative correlation with model predictions. Colors indicate the numerical value of the feature (red=high, blue=low).

TABLE III
DATA SUMMARIZING THE PAIRED STUDENT'S t-TEST BETWEEN THE PREDICTED DIFFERENCES BETWEEN MODELS A AND B. WE REPORT THE MEANS OF THE PAIRED DIFFERENCES, THE t-STATISTIC, p-VALUES, AND CONFIDENCE INTERVALS WITH A 95% CONFIDENCE LEVEL. WE OBSERVE STATISTICAL SIGNIFICANCE FOR EACH PROPERTY EXCEPT STRAIN, ASSUMING A SIGNIFICANCE LEVEL OF 5%.

	$A-B$ mean	t-statistic	p-value	CI-low	CI-high
Strength	4.50E-03	1.08E+01	2.00E-26	3.68E-03	5.31E-03
Strain	-5.98E-04	-8.31E-01	4.06E-01	-2.01E-03	8.14E-04
Toughness	1.45E-03	2.36E+00	1.82E-02	2.47E-04	2.65E-03
Diameter	1.28E-03	4.77E+00	1.97E-06	7.51E-04	1.80E-03
Young's modulus	7.86E-04	2.31E+00	2.10E-02	1.18E-04	1.45E-03

properties, we get the best performance around $K = 5$ features. Additional features seem to either not add additional information or even worsen performance. We speculate that this decrease could be that the additional features are correlated with the previous, thus adding unnecessary complexity while not providing additional structure. For the diameter, we see a constant performance gain by increasing K. For Young's modulus, increasing K does not seem to increase the performance much. For all features, the RMSE for models with only the top feature ($K = 1$), NaCl, provides nearly identical RMSE compared to a baseline model only predicting the mean of the property, whose RMSE is equal to the (biased) sample standard deviation (see Table II for the unbiased sample SD). For Young's modulus, our model was not able to capture more complexity than the sample variance, as evidenced by $R^2 \approx 0$.

V. CONCLUSION

We developed machine learning models to predict the mechanical properties of recombinant spider silk fibers from spinning conditions, with and without pretrained protein embeddings. Both models captured overall trends, with diameter predicted most accurately and Young's modulus remaining difficult due to noise and low variance. Embeddings provided small but significant gains, suggesting sequence contributes alongside experimental factors. Feature analyses identified the amount of NaCl, spinning temperature, and protein type as key drivers, with a few features explaining most variance. These results show that while ML can guide silk optimization, progress will require larger datasets, improved measurements, and richer embeddings.

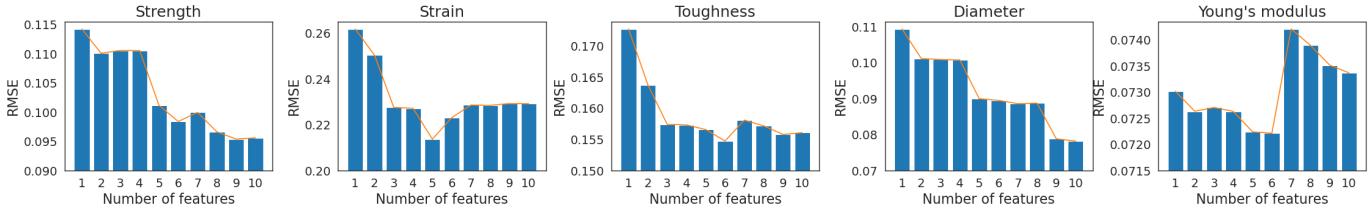

Fig. 4. RMSE for models trained by only including the top $K \in \{1, 2, ..., 10\}$ features derived from the average SHAP-ranking for all properties. We see an overall decrease increasing K, except for Young's modulus, where additional features do not meaningfully model performance.

REFERENCES

[1] L. Shen, E. Worrell, and M. Patel, "Present and future development in plastics from biomass," *Biofuels, Bioproducts and Biorefining*, vol. 4, no. 1, pp. 25–40, 2010.

[2] F. Vollrath and D. P. Knight, "Liquid crystalline spinning of spider silk," *Nature*, vol. 410, pp. 541–548, 2001.

[3] A. Rising, M. Widhe, J. Johansson, and M. Hedhammar, "Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications," *Cellular and Molecular Life Sciences*, vol. 68, no. 2, pp. 169–184, 2011.

[4] R. V. Lewis, "Spider silk: Ancient ideas for new biomaterials," *Chemical Reviews*, vol. 106, no. 9, pp. 3762–3774, 2006.

[5] M. Xu and R. V. Lewis, "Structure of a protein superfiber: spider dragline silk," *Proceedings of the National Academy of Sciences*, vol. 87, no. 18, pp. 7120–7124, 1990.

[6] A. Rising and J. Johansson, "Toward spinning artificial spider silk," *Nature Chemical Biology*, vol. 11, no. 5, pp. 309–315, 2015.

[7] M. Andersson, Q. Jia, A. Abella, X. Y. Lee, M. Landreh, P. Purhonen, H. Hebert, M. Tenje, C. V. Robinson, Q. Meng, G. R. Plaza, J. Johansson, and A. Rising, "Biomimetic spinning of artificial spider silk from a chimeric minispidroin," *Nature Chemical Biology*, vol. 13, no. 3, pp. 262–264, 2017.

[8] A. Koeppel and C. Holland, "Progress and trends in artificial silk spinning: A systematic review," *ACS Biomaterials Science & Engineering*, vol. 3, no. 3, pp. 226–237, 2017.

[9] V. G. Debabov and V. G. Bogush, "Recombinant spidroins as the basis for new materials," *ACS biomaterials science & engineering*, vol. 6, no. 7, pp. 3745–3761, 2020.

[10] M. Heim, D. Keerl, and T. Scheibel, "Spider silk: From soluble protein to extraordinary fiber," *Angewandte Chemie International Edition*, vol. 48, no. 20, pp. 3584–3596, 2009.

[11] A. Rising and collaborators, "Experimental dataset on recombinant spider silk fibers," unpublished dataset, Swedish University of Agricultural Sciences (SLU).

[12] B. Schmuck, G. Greco, F. G. Bäcklund, N. M. Pugno, J. Johansson, and A. Rising, "Impact of physio-chemical spinning conditions on the mechanical properties of biomimetic spider silk fibers," *Communications Materials*, vol. 3, no. 83, 2022. [Online]. Available: https://doi.org/10.1038/s43246-022-00307-6

[13] W. Lu, D. L. Kaplan, and M. J. Buehler, "Generative modeling, design, and analysis of spider silk protein sequences for enhanced mechanical properties," *Advanced Functional Materials*, vol. 34, no. 11, p. 2311324, 2024. [Online]. Available: https://doi.org/10.1002/adfm.202311324

[14] N. Dubey, E. Karlsson, M. A. Redondo, J. Reimegård, A. Rising, and H. Kjellström, "Customizing spider silk: Generative models with mechanical property conditioning for protein engineering," *arXiv preprint arXiv:2504.08437*, April 2025. [Online]. Available: https://arxiv.org/abs/2504.08437

[15] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, T. Gibbs, T. Feher, C. Angerer, M. Steinegger, D. Bhowmik, and B. Rost, "Prottrans: Towards cracking the language of life's code through self-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7112–7127, 2021.

[16] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, R. Fergus *et al.*, "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," *Proceedings of the National Academy of Sciences*, vol. 118, no. 15, p. e2016239118, 2021.

[17] R. J. Young, C. Holland, Z. Shao, and F. Vollrath, "Spinning conditions affect structure and properties of nephila spider silk," *MRS Bulletin*, vol. 46, no. 10, pp. 915–924, 2021. [Online]. Available: https://doi.org/10.1557/s43577-021-00194-1

[18] J. Pérez-Rigueiro, M. Elices, G. R. Plaza, J. I. Real, and G. V. Guinea, "The effect of spinning forces on spider silk properties," *Journal of Experimental Biology*, vol. 208, no. 14, pp. 2633–2639, 2005.

[19] J. Zhang, M. Gong, and Q. Meng, "Wet spinning is employed to produce spider silk with high elasticity," *APL Materials*, vol. 11, no. 8, p. 081110, 2023. [Online]. Available: https://doi.org/10.1063/5.0160351

[20] G. Greco, B. Schmuck, S. K. Jalali, N. M. Pugno, and A. Rising, "Influence of experimental methods on the mechanical properties of silk fibers: A systematic literature review and future road map," *Biophysics Reviews*, vol. 4, no. 3, p. 031301, 2023.

[21] Y. Wang, J. Li, and M. J. Buehler, "Generative design and molecular mechanics characterization of silk proteins," *Materials Advances*, vol. 6, no. 12, pp. 1543–1552, 2025. [Online]. Available: https://doi.org/10.1039/d5ma00154d

[22] B. Schmuck, G. Greco, T. B. Pessatti, S. Sonavane, V. Langwallner, T. Arndt, and A. Rising, "Strategies for making high-performance artificial spider silk fibers," *Advanced Functional Materials*, vol. 34, no. 35, p. 2305040, 2024.

[23] T. Arndt, G. Greco, B. Schmuck, J. Bunz, O. Shilkova, J. Francis, N. M. Pugno, K. Jaudzems, A. Barth, J. Johansson, and A. Rising, "Engineered spider silk proteins for biomimetic spinning of fibers with toughness equal to dragline silks," *Advanced Functional Materials*, vol. 32, no. 23, p. 2200986, 2022.

[24] B. Schmuck, G. Greco, A. Barth, N. M. Pugno, J. Johansson, and A. Rising, "High-yield production of a super-soluble miniature spidroin for biomimetic high-performance materials," *Materials Today*, vol. 50, pp. 16–23, 2021.

[25] A. Rising and J. Johansson, "Toward spinning artificial spider silk," *Nature Chemical Biology*, vol. 11, no. 5, pp. 309–315, 2015.

[26] T. Arndt, P. R. Laity, J. Johansson, C. Holland, and A. Rising, "Native-like flow properties of an artificial spider silk dope," *ACS Biomaterials Science & Engineering*, vol. 7, no. 2, pp. 462–471, 2021.

[27] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli *et al.*, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, no. 6637, pp. 1123–1130, 2023.

[28] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[29] D. McElfresh, S. Khandagale, J. Valverde, V. Prasad C, G. Ramakrishnan, M. Goldblum, and C. White, "When do neural nets outperform boosted trees on tabular data?" *Advances in Neural Information Processing Systems*, vol. 36, pp. 76 336–76 369, 2023.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[31] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[32] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.