Department of Electrical and Computer Engineering
North South University

Physics-Guided Machine Learning for Accurate GSM Prediction in Single Jersey Knit Fabric: A Comprehensive Comparative Study with Explainable AI

Md. Aziz Raihan	ID 2132681042
Amanullah Ahsan	ID 2021769642
Amiur Rahman	ID 2131466642
Md. Mushfiqur Rashid Marmo	ID 2021496642

Faculty Advisor:
Dr. Sifat Momen
Professor
ECE Department

Fall, 2025

ABSTRACT

Physics-Guided Machine Learning for Accurate GSM Prediction in Single Jersey Knit Fabric: A Comprehensive Comparative Study with Explainable AI

Accurate prediction of fabric weight (GSM, grams per square meter) is essential for quality control, cost estimation, and production planning in textile manufacturing. Conventional approaches rely on empirical formulas or post-production measurements, limiting predictive capability during manufacturing. This study presents a comprehensive machine learning framework for GSM prediction in Single Jersey knit fabric by integrating physics-guided feature engineering with explainable artificial intelligence techniques. Using a large-scale production dataset collected from Fakhruddin Textile Mills, Bangladesh, this work focuses on 5,648 Single Jersey (S/J) samples selected from a broader multi-construction dataset due to sufficient sample size and complete feature availability.

A wide range of regression models was systematically evaluated, including interpretable linear and polynomial models, decision trees, generalized additive models, and ensemble-based approaches. Among the interpretable models, Polynomial Ridge regression (degree 3) achieved the best performance, with a test RMSE of approximately 7.35 $\mathrm{g\,m^{-2}}$ and test $R^2 \approx 0.845$, demonstrating the benefit of nonlinear modeling while retaining transparency. However, ensemble-based models significantly outperformed interpretable approaches. In particular, the Extra Trees ensemble achieved the best overall test performance, with a test RMSE of approximately 4.10 $\mathrm{g\,m^{-2}}$, test $R^2 \approx 0.95$, and test MAPE below 1%.

Although individual feature to GSM correlations are modest (maximum $r = 0.328$), physics-guided feature engineering, especially the inclusion of the tightness factor, proved effective in improving predictive accuracy. SHAP-based explainability confirms that yarn count, tightness factor, and stitch length dominate model predictions, together accounting for more than 86% of total feature importance, consistent with established textile engineering principles. Cross-validation analysis further demonstrates stable generalization, with the Extra Trees model achieving a cross-validated RMSE of 2.81 ± 0.47 $\mathrm{g\,m^{-2}}$.

From an industrial perspective, the achieved reduction in GSM prediction error supports tighter specification control and improved process consistency. Under conservative assumptions commonly reported in textile manufacturing practice, such improvements may correspond to indicative annual cost savings on the order of \$100k to \$200k for medium-sized mills, although detailed economic evaluation remains an area for future work. Overall, the results demonstrate that combining textile physics with nonlinear ensemble learning enables accurate, stable, and interpretable GSM prediction suitable for data-driven quality monitoring and process optimization.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background and Motivation

Textile manufacturing is a complex, multi-parameter process where machine settings, material properties, and operational conditions collectively determine final fabric characteristics. Among these characteristics, Grams per Square Meter (GSM) the fabric weight per unit area that stands as one of the most critical quality indicators. GSM directly governs material consumption, production cost, buyer acceptance criteria, and overall quality assurance in textile manufacturing [18, 27].

The precision of GSM control carries profound economic significance, particularly in countries where the textile sector forms a cornerstone of the national economy. Bangladesh exemplifies this dependency: in fiscal year 2023–2024, the Ready-Made Garment (RMG) sector achieved export revenues of USD 47.38 billion, constituting approximately 82% of the nation's total export earnings and employing over 4 million workers [6]. Within this landscape, any inefficiency in quality control including GSM variation that translates into substantial economic losses at both enterprise and national scales. Conservative industry estimates suggest that GSM-related defects account for 5–10% of fabric rejection in knitting mills, leading to material wastage, production delays, and penalty costs [21, 26].

From a manufacturing perspective, inadequate GSM prediction results in excessive yarn consumption, elevated fabric rejection rates, rework costs, and reduced production throughput. Buyers impose stringent GSM tolerance specifications (typically ±3–5%), and deviations beyond these limits frequently result in order rejection, price penalties, or contractual disputes [28]. Despite its importance, GSM measurement in conventional practice occurs post-production through manual sampling and laboratory testing, rendering the process reactive, time-consuming, labor-intensive, and incapable of supporting real-time process adjustments [22].

Traditional GSM estimation methods rely predominantly on two approaches: empirical formulas derived from Pierce's geometric principles [32] and statistical regression models based on historical production data [21]. While these methods provide theoretical foundations, they suffer from critical limitations. Empirical formulas assume idealized yarn and fabric geometries, neglecting real-world complexities such as yarn irregularities, relaxation phenomena, and machine-induced variations. Statistical regression models, when restricted to linear formulations, fail to capture the inherently nonlinear relationships between input parameters and fabric weight [27]. Consequently, prediction errors often exceed acceptable industrial tolerances, limiting their utility for proactive quality control.

The global textile industry is undergoing a transformative shift toward Industry 4.0, characterized by data-driven automation, intelligent process monitoring, and predictive quality control systems [31]. In this context, machine learning (ML) techniques offer a paradigmatic alternative to conventional methods by enabling faster, more accurate, and scalable GSM estimation directly from production-stage parameters [5, 34]. Recent studies have demonstrated that ensemble learning methods, neural networks, and gradient boosting algorithms can effectively model complex, nonlinear relationships in textile processes [9, 20].

However, existing machine learning approaches to GSM prediction exhibit notable limitations. First, prior work predominantly evaluates isolated models without system-

atic benchmarking across interpretable and black-box approaches, limiting comparative insights [11, 26]. Second, hyperparameter optimization-critical for achieving robust generalization, is often inadequately documented or entirely omitted [34]. Third, the absence of physics-based feature engineering results in models that may achieve high predictive accuracy through spurious correlations rather than capturing fundamental textile mechanics [5]. Fourth, explainability analysis is rarely conducted, leaving model predictions opaque and difficult to validate against domain knowledge [1].

This study addresses these limitations by proposing a comprehensive physics-guided machine learning framework for GSM prediction in Single Jersey knit fabric. Single Jersey (S/J) is selected as the focal construction due to its industrial significance constituting over 47% of Bangladesh's knit fabric production and its dominance in the dataset (5,648 samples, 44.9%) [6]. The framework integrates Pierce's tightness factor [32], a theoretically grounded parameter relating yarn linear density and stitch length, as a physics-guided feature alongside conventional machine and yarn parameters. By systematically evaluating 22 regression models spanning interpretable (linear, polynomial, decision tree) and black-box (ensemble, boosting, neural network) approaches, applying rigorous hyperparameter optimization, conducting cross-validation stability analysis, and employing SHAP-based explainability [25], this work establishes a transparent, reproducible, and industrially deployable GSM prediction pipeline.

The primary motivation for this research is twofold: *scientific rigor* and *industrial applicability*. Scientifically, the study investigates whether explicitly encoding textile physics into feature representations enhances model performance and interpretability compared to purely empirical data-driven methods. Industrially, accurate GSM prediction enables proactive quality control, reduces material waste, minimizes rejection rates, and supports real-time decision-making in manufacturing environments transitioning toward data-driven operations. By bridging domain knowledge with state-of-the-art machine learning techniques, this work contributes to the emerging field of physics-informed machine learning for textile engineering [31].

1.2 Purpose and Goal of the Project

The primary purpose of this project is to develop accurate and interpretable machine learning models to predict the GSM of Single Jersey knit fabrics using production-stage parameters. The central research question addressed in this study is the following.

Can interpretable machine learning models be designed to accurately predict GSM while capturing the underlying physics of textile structures?

To answer this question, the project pursues the following key goals.

- **Physics-Guided Feature Engineering:**
  Incorporate domain knowledge from textile engineering specifically the tightness factor $K = \frac{\sqrt{\text{tex}}}{\text{stitch length}}$ derived from Pierce's equations—to enhance model performance and interpretability.

- **Comprehensive Model Benchmarking:**
  Evaluate a wide range of machine learning algorithms, including interpretable models (linear regression, generalized additive models, decision trees) and black-box models (ensemble methods and gradient boosting), to establish reliable performance baselines.

- **Accuracy–Interpretability Trade-off Analysis:**
  Design a dual-model framework that supports industrial deployment by balancing prediction accuracy and interpretability, enabling informed decision-making in real manufacturing environments.

- **Explainability and Trustworthiness:**
  Apply explainable AI techniques such as SHAP and LIME to verify that model predictions align with established textile physics, ensuring transparency and industrial acceptance.

- **Production-Ready Implementation:**
  Develop a complete prediction pipeline including hyperparameter optimization, cross-validation stability analysis, and a user-friendly prediction interface suitable for real-world deployment.

1.3 Organization of the Report

This report is organized into five chapters to systematically present the background, methodology, experimental evaluation, and conclusions of the proposed physics-guided machine learning framework for GSM prediction in Single Jersey knit fabric.

Chapter 1 introduces the research problem and its industrial significance. It presents the background and motivation for accurate GSM prediction in textile manufacturing and outlines the objectives and scope of the project.

Chapter 2 reviews relevant literature related to GSM prediction techniques, including traditional empirical and statistical models, soft computing approaches, and modern machine learning methods. This chapter also identifies key limitations in existing studies and highlights the research gaps addressed in this work.

Chapter 3 describes the methodology adopted in this study. It details the dataset characteristics, physics-guided feature engineering using the tightness factor, data preprocessing procedures, model selection strategy, hyperparameter optimization process, evaluation metrics, and explainable AI techniques used for model validation.

Chapter 4 presents the experimental results, analysis, and discussion. It compares the performance of multiple machine learning models, analyzes the accuracy–interpretability trade-off, evaluates model stability and robustness, and interprets the predictions using SHAP-based explainability in the context of textile physics.

Chapter 5 concludes the report by summarizing the key findings and contributions of the project. It also discusses the limitations of the current study and outlines possible directions for future improvements and extended research in physics-guided machine learning for textile manufacturing.

Chapter 2
Literature Review

The prediction of fabric weight (GSM) has evolved from empirical formulas rooted in textile geometry to sophisticated data-driven approaches leveraging machine learning. This section critically examines the trajectory of GSM prediction research, evaluates methodological advances and persistent limitations, and positions our physics-guided ensemble learning framework within the current state-of-the-art.

2.1 From Geometric Theory to Statistical Modeling

The theoretical foundation of fabric weight prediction traces to Pierce's pioneering work [32], which established mathematical relationships between yarn linear density, stitch geometry, and fabric dimensions. Pierce introduced the tightness factor $K = \sqrt{T}/l$ (where T is yarn tex and l is stitch length) as a dimensionless parameter characterizing loop compactness, a concept that remains fundamental to textile engineering seven decades later [18]. However, Pierce's geometric models assumed idealized conditions: perfectly circular yarn cross-sections, uniform stitch formation, and complete fabric relaxation. These assumptions are rarely satisfied in industrial production [36].

Subsequent theoretical refinements incorporating yarn flattening and relaxation phenomena achieved marginal improvements but introduced substantial computational complexity and construction-specific parameterization requirements [23]. These physics-based models, while intellectually rigorous, failed to achieve practical adoption in manufacturing environments due to their sensitivity to material variations and limited generalization across diverse production conditions [27].

The computational accessibility of statistical regression prompted a methodological shift toward data-driven modeling. Early studies employed ordinary least squares regression with yarn count, stitch length, and machine gauge as predictors, achieving modest explanatory power (R^2 approximately 0.75 to 0.80) [21]. Malik and Usman [28] demonstrated that polynomial regression captured some nonlinearities but suffered from extrapolation instability beyond training ranges. A critical limitation of purely statistical approaches (consistently overlooked in prior work) is their inability to encode domain knowledge explicitly, risking spurious correlations that lack physical justification [34].

2.2 Computer Vision and Automated Feature Extraction

Digital imaging technologies enabled automated fabric characterization through computer vision pipelines that extract structural features (yarn diameter, stitch density, loop dimensions) for subsequent regression modeling [22]. While achieving promising accuracy (MAPE below 5%) for plain constructions under controlled conditions, these methods exhibit critical operational limitations: sensitivity to lighting variations, surface texture dependence, and inability to handle complex multi-component structures [17, 40]. More fundamentally, image-based approaches predict GSM from finished fabric samples rather than production-stage parameters, precluding proactive quality control [29].

2.3 Neural Networks and Early Machine Learning

Artificial neural networks emerged as data-driven alternatives to rule-based systems. Behera and Mishra [5] pioneered multilayer perceptrons for predicting knitted fabric properties, achieving R^2 greater than 0.90. Majumdar et al. [26] confirmed that ANNs consistently outperformed linear regression but revealed vulnerability to overfitting on small, noisy industrial datasets, a persistent challenge given typical data availability in textile manufacturing [30]. Despite predictive success, early neural networks suffered from the "black-box" problem: network weights provide minimal insight into physical relationships, limiting interpretability essential for industrial adoption [15].

Recent systematic reviews document an exponential increase in textile AI publications, with 44% of papers published between 2022 and 2024 [42], yet comprehensive benchmarking across model architectures remains scarce. Metin and Bilgin [29] evaluated AutoML frameworks for fabric quality prediction, finding Random Forests superior to neural networks for structured tabular data, a pattern consistent with broader manufacturing applications [8].

2.4 The Rise of Ensemble Learning

The maturation of ensemble methods (Random Forests [7], Gradient Boosting [13], XGBoost [9], and LightGBM [20]) introduced algorithms combining multiple weak learners to achieve superior predictive performance with implicit regularization. Ramesh et al. [34] demonstrated that Random Forests reduced RMSE by 15 to 20% compared to ANNs for knitted fabric prediction through bootstrap aggregating and random feature sampling. However, systematic application to GSM prediction with rigorous hyperparameter optimization is conspicuously absent from textile literature [3].

Critically, existing textile ML studies exhibit methodological deficiencies. First, isolated model evaluations without comprehensive benchmarking [26]. Second, training with default hyperparameters rather than systematic optimization [11]. Third, absence of cross-validation stability analysis [34]. These omissions raise concerns about reproducibility and generalization claims.

2.5 Physics-Informed Machine Learning: An Emerging Paradigm

A transformative development in scientific machine learning, physics-informed machine learning (PIML), integrates domain knowledge directly into model architectures, training procedures, or feature representations [19]. Rather than treating ML as purely data-driven, PIML encodes governing equations, conservation laws, or theoretical relationships as constraints [39]. Recent reviews document PIML's success in additive manufacturing for temperature field prediction [12, 38], melt pool dynamics [41], and process-structure-property modeling [37]. These are domains where both data and domain knowledge coexist.

PIML offers three key advantages. First, enhanced generalization with limited data by leveraging physical constraints [19]. Second, improved interpretability through physics-consistent learned relationships [12]. Third, robustness to out-of-distribution scenarios through mechanistic grounding [37]. Physics-Informed Neural Networks (PINNs), which embed differential equations directly into loss functions, achieve remarkable data effi-

ciency by reducing training requirements by orders of magnitude compared to pure data-driven approaches [16, 33].

Despite theoretical appeal and manufacturing success, PIML remains underexplored in textile engineering. The opportunity to integrate Pierce's tightness factor (a well-established physics-based feature) into ML frameworks represents a significant research gap, particularly for manufacturing prediction tasks where both production data and domain theory are available [42].

2.6 Explainable AI in Manufacturing

The "black-box" nature of modern ML models has driven demand for explainable AI (XAI) techniques, ensuring transparency and trustworthiness [2]. Two prominent frameworks, LIME (Local Interpretable Model-agnostic Explanations) [35] and SHAP (SHapley Additive exPlanations) [25], provide post-hoc interpretability by quantifying feature contributions to predictions.

SHAP, grounded in cooperative game theory, assigns each feature a Shapley value representing its marginal contribution. It satisfies desirable theoretical properties (local accuracy, missingness, consistency) that LIME does not guarantee [25]. Recent manufacturing applications demonstrate SHAP's efficacy: semiconductor yield prediction [24], quality control in laser powder bed fusion [14], and fault diagnosis [10]. Critically, SHAP enables validation that learned feature importances align with domain knowledge, a capability essential for establishing model trustworthiness beyond performance metrics [2].

Despite growing adoption in manufacturing [2,10], XAI applications to textile property prediction remain minimal. To our knowledge, no prior work has applied SHAP or LIME to GSM prediction to validate physics consistency. This gap undermines confidence in ML predictions for industrial deployment.

2.7 Research Gaps and Study Positioning

Comprehensive literature analysis reveals five critical gaps motivating this work:

G1 (Absence of Comprehensive Benchmarking): Prior studies evaluate isolated models or pairwise comparisons without systematic assessment across interpretable and ensemble approaches, cross-validation, and stability analysis [26, 34].

G2 (Inadequate Hyperparameter Optimization): Models are trained with default library settings; systematic optimization (grid search, Bayesian optimization) is rarely documented, limiting reproducibility and potentially yielding suboptimal performance claims [11, 29].

G3 (Lack of Physics-Guided Feature Engineering): Despite well-established textile theory, ML models rely exclusively on raw parameters without incorporating derived physics-based features. Pierce's tightness factor, which directly relates yarn and stitch properties to structure, is absent from existing ML feature sets [5, 26, 42].

G4 (Absence of Explainability and Physics Validation): Ensemble models are rarely subjected to SHAP/LIME analysis to validate that learned patterns correspond to textile physics. Without such validation, models may achieve high accuracy through spurious correlations rather than genuine relationships [2].

G5 (Construction-Specific vs. Generalized Modeling Ambiguity): Literature inconsistently combines heterogeneous fabric types in unified datasets, potentially

introducing confounding structural effects, or focuses on single constructions without justification or generalization discussion [3, 27].

This study addresses these gaps through a comprehensive physics-guided ensemble learning framework. Specifically, our contributions include: (1) systematic benchmarking of 22 algorithms spanning interpretable and black-box approaches with rigorous cross-validation, (2) RandomizedSearchCV hyperparameter optimization with reproducible documentation, (3) explicit incorporation of Pierce's tightness factor with ablation validation, (4) SHAP/LIME analysis to validate physics consistency, and (5) construction-specific modeling (Single Jersey) with explicit methodological justification.

By integrating domain knowledge with modern ensemble learning and rigorous explainability analysis, this work advances the state-of-the-art in physics-guided machine learning for textile manufacturing, providing accurate predictions and interpretable insights that support industrial adoption.

Chapter 3
Methodology

This chapter presents the complete methodological framework for predicting GSM in Single Jersey knit fabric through physics-guided machine learning. We describe our systematic approach progressing from raw production data to validated predictive models, emphasizing reproducibility and physical consistency throughout the pipeline. Figure 1 provides an overview of the complete methodology.

3.1 Methodology Pipeline

Our approach consists of eight interconnected stages designed to balance predictive accuracy with interpretability and physical validity. The pipeline begins with comprehensive data quality control, where we filter the complete dataset to retain only Single Jersey constructions, ensuring structural homogeneity. This focused approach avoids the confounding effects that would arise from combining fundamentally different fabric architectures.

Feature engineering integrates direct measurements with physics-derived variables, explicitly encoding textile engineering knowledge into the model inputs. We then partition the data into training, validation, and test subsets following rigorous protocols to prevent information leakage and ensure honest performance evaluation.

Model selection spans a diverse range of algorithms, from simple linear regression to complex ensemble methods, establishing comprehensive performance baselines across the interpretability-accuracy spectrum. The top-performing models undergo systematic hyperparameter optimization using randomized search with cross-validation. Beyond baseline evaluation, we explore advanced techniques including Generalized Additive Models and meta-ensemble strategies.

Stability assessment through multiple cross-validation runs and different random seeds ensures our results reflect genuine model capability rather than fortunate data partitioning. Performance evaluation employs multiple complementary metrics, each capturing different aspects of prediction quality relevant to industrial deployment.

The pipeline culminates in explainability analysis, where we validate that learned patterns align with established textile physics rather than spurious correlations. This comprehensive approach ensures the final model achieves strong predictive performance while maintaining the transparency and physical consistency required for industrial acceptance.

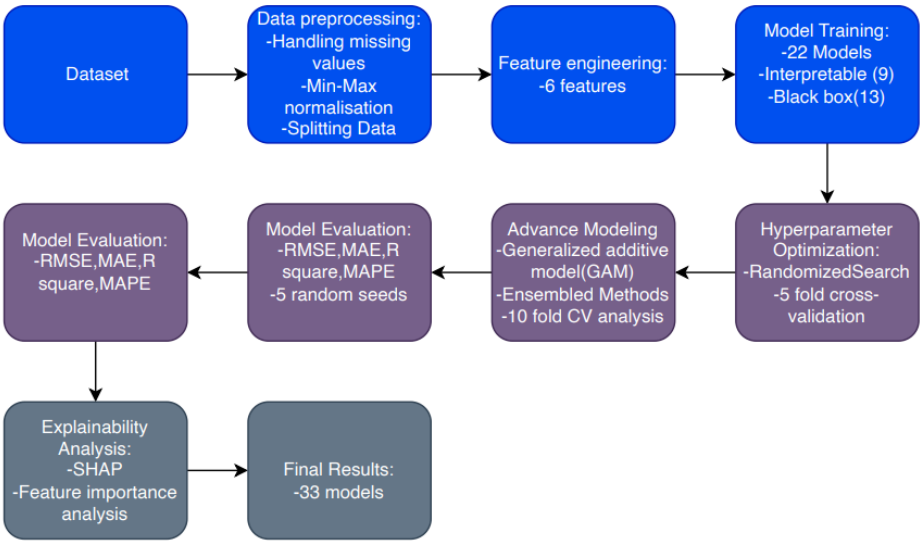


Figure 1: Methodology block diagram

3.2 Dataset Description

The dataset used in this study is derived from the structured knitting dataset introduced by Ahmed and Junayed [4], which was developed specifically to support machine learning and data mining applications in textile manufacturing. The raw production data were collected from an industrial knitting facility and subsequently processed using domain-informed preprocessing techniques, including data cleaning, normalization, and physics-guided feature engineering. The resulting dataset provides a comprehensive and standardized representation of knitting process parameters relevant to GSM prediction.

The finalized dataset consists of 12,569 fabric samples with 38 recorded features, encompassing yarn properties, machine specifications, stitch characteristics, and derived physical parameters such as the tightness factor. The dataset includes eight distinct fabric construction types, reflecting realistic industrial production diversity. Single Jersey represents the largest subset with 5,648 samples, accounting for approximately 44.9% of the total dataset. The remaining samples correspond to Long Sleeve Jersey (1,970 samples, 15.7%), Rib (1,913 samples, 15.2%), Fleece Terry (1,168 samples, 9.3%), Mesh (621 samples, 4.9%), Piqué (571 samples, 4.5%), Interlock (571 samples, 4.5%), and a small number of miscellaneous constructions.

In this work, modeling is restricted to the Single Jersey subset for several reasons. First, its substantially larger sample size ensures statistically reliable training, validation, and testing, enabling robust learning of nonlinear relationships without compromising evaluation integrity. Second, the dataset paper emphasizes that higher-order yarn and stitch features are most consistently available for Single Jersey fabrics, whereas other constructions exhibit significant feature sparsity [4]. Third, different knit constructions possess fundamentally distinct loop geometries and mass distribution mechanisms, making unified modeling across constructions prone to structural confounding. For example, Rib fabrics incorporate alternating knit and purl loops, while Fleece Terry introduces pile loops, both of which alter GSM behavior relative to the uniform loop structure of Single Jersey fabrics.

The measured GSM values for Single Jersey samples range from approximately 115

to 260 g/m^2, with a mean of 152.8 g/m^2 and a standard deviation of 17.7 g/m^2. This wide and continuous distribution reflects the diversity of industrial products, spanning lightweight fabrics to heavier constructions. As noted in the dataset paper, prior exploratory analysis also revealed meaningful nonlinear relationships between GSM and key parameters such as yarn count and tightness factor, further motivating the suitability of Single Jersey fabrics for regression-based modeling [4]. Overall, the dataset provides a well-structured, physically grounded, and industrially representative foundation for developing and evaluating machine learning models for GSM prediction.

3.3 Physics-Guided Feature Engineering

Rather than treating GSM prediction as a purely empirical problem, we integrate established textile engineering principles through careful feature design. This physics-guided approach distinguishes our work from purely data-driven methods that rely only on raw measurements.

We selected six features combining direct measurements with derived physics-based variables. The five measured parameters come directly from production records. Yarn linear density (yarn_count_tex_1) is measured in tex units, where one tex equals one gram per 1,000 meters. This fundamental material property determines the mass of yarn in each loop. Stitch length (stitch_1) is measured in millimeters and controls loop size, affecting both fabric openness and weight. Machine diameter (m/c_dia) is specified in inches and determines the circumference of the fabric tube being produced. Machine gauge (m/c_gauge) counts needles per inch, establishing stitch density. The number of feeders (number_of_feeders) indicates how many yarn feeding positions are active during production, affecting fabric uniformity and production speed.

From these measurements, we compute the tightness factor using Pierce's classical formula established in 1947: $K = \sqrt{\text{tex}}/\text{stitch}$. This dimensionless quantity relates yarn thickness to loop size, capturing fundamental fabric geometry. Higher tightness factors indicate tighter structures where yarn diameter approaches stitch length, producing denser fabrics with higher GSM. Lower values indicate more open structures where loops are large relative to yarn thickness.

The tightness factor embeds decades of textile engineering knowledge into a single derived feature. Rather than requiring the model to learn this relationship from data alone, we provide it explicitly, reducing the complexity of patterns the model must discover. This approach aligns with the broader principle that incorporating domain knowledge improves both model performance and interpretability.

Correlation analysis reveals the relationships between features and the target GSM. Among individual features, yarn count shows the strongest linear correlation with GSM ($r = 0.328$, $p < 0.001$), confirming that yarn thickness directly influences fabric weight. Stitch length exhibits a positive correlation ($r = 0.180$, $p < 0.001$), where longer stitches generally associate with heavier fabrics in this dataset. The tightness factor shows weaker linear correlation ($r = 0.138$, $p < 0.001$) than might be initially expected from textile theory.

This modest correlation for the tightness factor reflects an important distinction. While textile theory predicts that the tightness factor affects GSM, the relationship is nonlinear and interacts with other parameters. The relatively weak linear correlation does not invalidate the physics; rather, it indicates that simple linear models will struggle to leverage this feature effectively. More sophisticated nonlinear models can potentially

extract greater predictive value from tightness factor through interactions with other variables. Our explainability analysis in later sections examines whether models successfully learn these physics-based relationships.

Table 1 summarizes the selected features, their units, and physical significance:

Feature	**Units**	**Physical Significance**
yarn_count_tex_1	tex	Primary yarn linear density, directly determines material mass
stitch_1	mm	Loop geometry and fabric openness
m/c_dia	inches	Fabric tube size and circumference
m/c_gauge	count	Needle density affecting stitch count per unit area
number_of_feeders	count	Yarn feeding positions influencing production uniformity
tightness_factor	dimensionless	Integrated yarn-stitch physics from Pierce's theory

Table I: Selected features for GSM prediction

3.4 Data Partitioning and Normalization

Proper data partitioning is critical for honest performance evaluation. We split the 5,648 Single Jersey samples into training (70%), validation (15%), and test (15%) subsets, yielding 3,952 training samples, 848 validation samples, and 848 test samples. The split proceeds in two stages: first separating 15% as the held-out test set, then dividing the remaining 85% into training and validation subsets using a 0.1765 split ratio (which produces exactly 15% of the original total for validation).

We fix the random seed at 42 throughout all experiments, ensuring complete reproducibility. The same samples appear in training, validation, and test sets across all model runs, enabling fair comparison between algorithms. The validation set guides model selection and hyperparameter tuning decisions, while the test set remains completely isolated until final evaluation, never influencing any training or tuning decisions.

This three-way split serves distinct purposes. The training set provides data for learning model parameters such as regression coefficients or tree structures. The validation set allows us to compare different model architectures and hyperparameter configurations without touching the test set. Only after all modeling decisions are finalized do we evaluate on the test set, which provides our best estimate of how models will perform on future production data.

Feature normalization applies StandardScaler transformation, which converts each feature to zero mean and unit variance using z-score normalization: $x_{scaled} = (x - \mu)/\sigma$. Critically, we compute the mean and standard deviation exclusively from the training set, then apply these same parameters to transform validation and test sets. This prevents information leakage where test set statistics would contaminate the training process.

Normalization proves essential for several model families in our ensemble. Distance-based methods such as support vector machines and K-Nearest Neighbors perform poorly when features exist on vastly different scales, as they would be dominated by large-

magnitude variables. Neural networks train more efficiently with normalized inputs, avoiding saturation in activation functions. Regularized linear models benefit from comparable feature scales when applying penalties.

However, tree-based algorithms operate differently. Decision trees and ensemble methods like Random Forests are invariant to monotonic feature transformations because they make decisions based on threshold comparisons rather than distances or gradients. For these models, we train on unnormalized data to avoid unnecessary preprocessing. Our implementation automatically selects normalized or unnormalized features based on the algorithm being trained.

3.5 Model Selection Strategy

We trained 22 diverse regression algorithms to understand how different modeling approaches handle textile manufacturing data. This breadth serves multiple purposes: establishing comprehensive performance baselines, quantifying tradeoffs between interpretability and accuracy, and avoiding the common pitfall of reporting only the best model while ignoring alternatives that might offer other advantages.

The models divide into interpretable and black-box categories based on their transparency to human inspection. Interpretable models provide explicit mathematical relationships that textile engineers can examine and validate.

Interpretable Models (9 algorithms):

- **Ordinary Least Squares (OLS):** Linear regression without regularization, serving as the simplest baseline.
- **Ridge Regression:** Adds L2 penalty to prevent coefficient magnitude from growing too large.
- **Lasso Regression:** Uses L1 penalty that can drive some coefficients exactly to zero, performing automatic feature selection.
- **ElasticNet:** Combines L1 and L2 penalties, balancing Lasso's feature selection with Ridge's stability.
- **Bayesian Ridge:** Treats coefficients as random variables with probabilistic priors, providing uncertainty estimates.
- **Huber Regressor:** Uses a robust loss function less sensitive to outliers than the squared error.
- **Polynomial Ridge (degree 2):** Expands features to include squared terms and interactions, then applies Ridge regularization.
- **Polynomial Ridge (degree 3):** Further expands to cubic terms, capturing more complex nonlinear relationships.
- **Decision Tree:** Learns hierarchical rules through recursive partitioning, limited to depth 5 to prevent overfitting.

Black-box models prioritize predictive accuracy over transparency. While their internal mechanisms are opaque, they can capture complex patterns that simpler models miss.

Black-Box Models (13 algorithms):

- **Random Forest:** Ensemble of 100 decision trees trained on bootstrap samples with random feature subsets.
- **Extra Trees:** Similar to Random Forest but uses random thresholds instead of optimal splits, increasing diversity.
- **Gradient Boosting:** Sequentially trains trees where each corrects errors of previous trees.
- **AdaBoost:** Adaptively reweights samples, focusing on difficult cases.
- **XGBoost:** Advanced gradient boosting with regularization and efficient implementation.
- **LightGBM:** Fast gradient boosting using histogram-based learning and leaf-wise growth.
- **CatBoost:** Gradient boosting with ordered target encoding and symmetric trees.
- **SVR-RBF:** Support vector regression with radial basis function kernel.
- **SVR-Poly:** Support vector regression with polynomial kernel (degree 2).
- **K-Nearest Neighbors (KNN):** Predicts based on average of 5 nearest training samples.
- **MLP-Small:** Neural network with single hidden layer (32 nodes).
- **MLP-Medium:** Neural network with two hidden layers (64 and 32 nodes).
- **MLP-Large:** Neural network with three hidden layers (128, 64, and 32 nodes).

Each model trains on the training set using default hyperparameters initially. We record training time, prediction time, and performance metrics on both validation and test sets. This baseline evaluation reveals which model families naturally align with textile data structure before investing effort in optimization.

Beyond these 22 baseline models, we investigated Generalized Additive Models (GAM) to explore the middle ground between interpretable and black-box approaches. GAM fits smooth nonlinear functions to individual features before combining them additively, potentially offering flexibility approaching neural networks while maintaining interpretability closer to linear models.

3.6 Hyperparameter Optimization

The six models showing the strongest baseline performance underwent systematic hyperparameter tuning. We selected Extra Trees, Random Forest, XGBoost, LightGBM, CatBoost, and SVR with RBF kernel based on their validation set RMSE. For each model, we defined parameter grids capturing key configuration choices.

We used RandomizedSearchCV to explore these parameter spaces efficiently. Unlike exhaustive grid search, which tests every possible combination, RandomizedSearchCV

samples 20 random configurations from each parameter grid. This approach balances thorough exploration against computational cost. Testing every combination would require hundreds of model fits per algorithm, while 20 samples provide good coverage of the parameter space at reasonable computational expense.

Each sampled configuration undergoes 5-fold cross-validation on the training set. The data splits into five parts, with the model trained on four parts and validated on the remaining part, cycling through all combinations. This produces five performance measurements for each configuration, whose average guides the selection of optimal parameters. We score configurations using negative mean squared error since scikit-learn's optimization framework treats all problems as minimization.

The parameter grids for each model reflect important algorithmic choices:

Model	**Hyperparameters and Search Space**	**Grid Size**
Extra Trees	*n_estimators*: 100, 200, 300; max_depth: None, 10, 20, 30; min_samples_split: 2, 5, 10; min_samples_leaf: 1, 2, 4; max_features: sqrt, log2, all	324
Random Forest	*n_estimators*: 100, 200, 300; max_depth: None, 10, 20, 30; min_samples_split: 2, 5, 10; min_samples_leaf: 1, 2, 4; max_features: sqrt, log2	216
XGBoost	*n_estimators*: 100, 200, 300; max_depth: 3, 5, 7, 10; learning_rate: 0.01, 0.05, 0.1, 0.2; subsample: 0.8, 0.9, 1.0; colsample_bytree: 0.8, 0.9, 1.0	432
LightGBM	*n_estimators*: 100, 200, 300; max_depth: 3, 5, 7, 10; learning_rate: 0.01, 0.05, 0.1, 0.2; num_leaves: 20, 31, 40, 50; subsample: 0.8, 0.9, 1.0	576
CatBoost	iterations: 100, 200, 300; depth: 4, 6, 8, 10; learning_rate: 0.01, 0.05, 0.1, 0.2; l2_leaf_reg: 1, 3, 5, 7	192
SVR (RBF)	C: 0.1, 1, 10, 100; gamma: scale, auto, 0.001, 0.01, 0.1; epsilon: 0.01, 0.1, 0.2, 0.5	80

Table II: Hyperparameter search spaces for all evaluated models. RandomizedSearchCV sampled 20 configurations per model.

3.7 Evaluation Metrics

We assessed models using four complementary metrics, recognizing that no single measure perfectly captures prediction quality for industrial applications.

Root Mean Square Error (RMSE) serves as our primary metric:

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

RMSE penalizes large errors more heavily than small ones due to the squaring operation. In textile manufacturing, extreme GSM deviations cause material waste, order rejection, and customer dissatisfaction. A fabric with GSM 20% above target might

be completely unusable, while minor deviations of 2-3% remain acceptable. RMSE's quadratic penalty aligns with these operational concerns by emphasizing accuracy on difficult cases.

Mean Absolute Error (MAE) provides a more robust alternative:

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

MAE measures typical error magnitude without amplifying outliers. When comparing MAE and RMSE, their relative values indicate whether errors concentrate around the average (similar MAE and RMSE) or whether occasional large errors dominate (RMSE substantially larger than MAE). This distinction helps assess whether model failures are systematic or sporadic.

Coefficient of Determination (R^2) indicates explained variance:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

R^2 ranges from negative infinity to 1, where 1 indicates perfect prediction, and 0 means the model performs no better than predicting the mean for every sample. Values approaching 1 confirm that the model captures most GSM variation. Negative values, though rare, indicate the model performs worse than the naive mean predictor, signaling fundamental problems.

Mean Absolute Percentage Error (MAPE) expresses accuracy as percentage:

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

MAPE provides intuitive interpretation for practitioners who naturally think in percentage terms. A MAPE of 2% means predictions typically deviate 2% from true values. However, MAPE can behave poorly when true values approach zero, so we use it alongside absolute metrics rather than exclusively.

All metrics compute on the held-out test set, which models never observe during training or hyperparameter selection. For models involving stochastic elements (ensemble methods, neural networks), we train multiple times with different random seeds and report both mean performance and standard deviation. This quantifies reproducibility and enables statistical comparison between models.

3.8 Explainability Analysis

Accurate predictions alone do not suffice for industrial adoption. Production engineers need to understand why models make specific predictions to build trust, diagnose failures, and validate consistency with established textile physics. We therefore apply explainability methods to examine model behavior, particularly for top-performing ensemble algorithms.

SHAP (SHapley Additive exPlanations) quantifies each feature's contribution to predictions through concepts from cooperative game theory. For any prediction, SHAP decomposes the difference between predicted GSM and average GSM into additive contributions from each feature. If a sample's predicted GSM is 160 g/m^2 and the dataset

average is 150 g/m², SHAP assigns each feature a value indicating how much it pushed the prediction above or below average. These feature attributions sum exactly to +10 g/m².

SHAP values satisfy important theoretical guarantees. They assign zero contribution to features that genuinely have no effect on the prediction. If two features contribute identically to a prediction, they receive equal SHAP values. When comparing different models, SHAP maintains consistency in how it attributes importance. These properties make SHAP particularly trustworthy compared to ad-hoc importance measures.

We compute SHAP values both globally and locally. Global SHAP analysis aggregates across all test samples, revealing which features drive predictions on average. We calculate the mean absolute SHAP value for each feature, quantifying its typical impact. Local SHAP analysis examines individual predictions, showing which features dominated that specific case. This helps us understand why the model made particular predictions and identify patterns in model behavior.

Beyond SHAP, we examine traditional feature importance scores from tree-based models. Random Forests and gradient boosting methods naturally produce importance rankings based on how much each feature reduces prediction error when included in splits. Comparing these rankings with SHAP importance validates consistency across different measurement approaches.

LIME (Local Interpretable Model-agnostic Explanations) provides an alternative perspective on individual predictions. LIME generates synthetic samples near a point of interest, obtains model predictions for those samples, then fits a simple linear model to approximate local behavior. This linear approximation may be easier to understand than global nonlinear relationships.

We apply LIME to selected test cases, particularly those with large prediction errors. Understanding why the model struggled on difficult cases reveals limitations and guides future improvements. We also examine cases with high confidence (low residuals) to verify that the model reasons correctly even when accurate.

The critical validation step checks whether learned feature importances align with textile engineering principles. If SHAP revealed that yarn color was more important than yarn count, we would suspect spurious correlations rather than genuine learning. If machine gauge-dominated predictions while yarn properties appeared irrelevant, the model might be exploiting confounding factors rather than causal relationships.

Instead, we expect yarn tex, stitch length, and tightness factor to drive predictions, consistent with fundamental physics of fabric formation. We examine whether features that correlate strongly with GSM also receive high SHAP importance, and whether the direction of effects matches physical intuition (for example, higher yarn count should increase predicted GSM). This validation distinguishes our work from studies reporting high accuracy without verifying that models learn meaningful patterns.

Chapter 4
Result, Analysis and Discussion

4.1 Results

A. Feature Correlation Analysis

Feature	Correlation (r)	p-value	Interpretation
tightness_factor	0.138	< 0.001	Weak positive correlation
yarn_count_tex_1	0.328	< 0.001	Strongest correlation
stitch_1	0.180	< 0.001	Longer stitch $\rightarrow$ higher GSM
m/c_gauge	-0.131	< 0.001	Weak negative correlation
number_of_feeders	-0.088	< 0.001	Negative correlation
m/c_dia	-0.236	< 0.001	Negative correlation

Table III: Feature–GSM correlation analysis

The apparent discrepancy between the two correlation values arises because they measure different types of relationships. The value 0.138 represents the linear correlation between *tightness_factor* and GSM, indicating that, when considered in isolation, tightness factor has only a weak direct linear association with fabric weight. This reflects the fact that GSM is influenced by multiple interacting variables rather than a single factor.

In contrast, the much higher correlation value of 0.939 corresponds to the relationship between *yarn_count_tex_1* and *tightness_factor*, which is expected because the tightness factor is a physics-derived feature that explicitly incorporates yarn count in its formulation. Since yarn count varies more widely than stitch length in the dataset, changes in yarn count dominate the variation in the tightness factor, leading to a strong inter-feature correlation.

Together, these observations highlight that high feature correlation does not imply strong feature target correlation. Instead, they emphasize the nonlinear and multivariate nature of GSM formation, justifying the use of physics-guided features in combination with nonlinear machine learning models rather than relying on simple linear relationships.

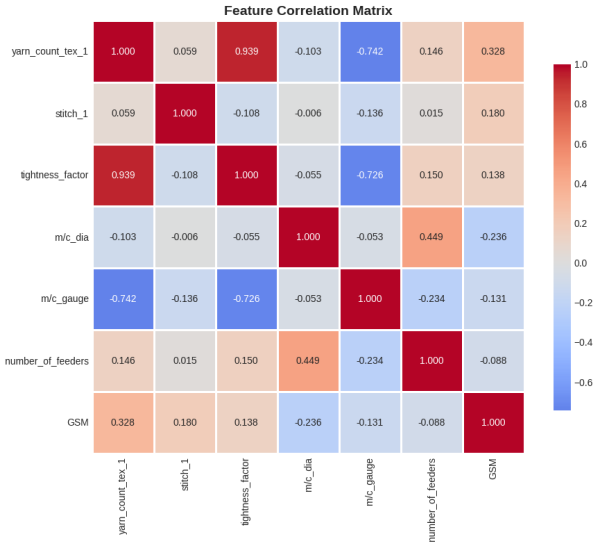


Figure 2: Feature correlation matrix

B. Model Performance

Model	Type	RMSE	MAE	R^2	MAPE
Extra Trees	Black-box	**4.10**	0.786	0.952	0.467
XGBoost	Black-box	4.68	0.937	0.937	0.560
LightGBM	Black-box	5.23	1.537	0.921	0.935
Random Forest	Black-box	5.36	0.981	0.917	0.572
KNN	Black-box	5.38	1.256	0.917	0.765
MLP-Medium	Black-box	5.81	2.933	0.903	1.855
Gradient Boosting	Black-box	5.92	2.009	0.899	1.266
MLP-Large	Black-box	5.94	2.669	0.899	1.683
Decision Tree	Interpretable	7.61	2.702	0.834	1.798
AdaBoost	Black-box	8.10	4.425	0.812	2.896
Poly-3 Ridge	Interpretable	**7.35**	4.457	0.845	2.892
SVR-RBF	Black-box	12.46	5.532	0.554	3.317
Ridge	Interpretable	15.05	10.128	0.349	6.414
Bayesian Ridge	Interpretable	15.05	10.129	0.349	6.415
OLS	Interpretable	15.05	10.305	0.349	6.416
Lasso	Interpretable	16.23	11.318	0.244	7.201
ElasticNet	Interpretable	17.24	12.454	0.146	7.941
Huber	Interpretable	17.58	7.983	0.112	4.743
SVR-Poly	Black-box	19.92	11.516	-0.140	7.252

Table IV: Test set performance comparison. Best-performing Black-box and Interpretable models (based on RMSE) are highlighted in bold.

The comparative evaluation of regression models on the test set reveals distinct performance patterns across different modeling approaches.

Among all evaluated models, **Extra Trees** achieved the best overall performance, recording the lowest test RMSE (4.10 g/m^2), the highest coefficient of determination ($R^2 = 0.952$), and the lowest MAPE (0.47%). This indicates its superior capability to capture complex, non-linear interactions among yarn count, stitch length, machine parameters, and other production features. As a result, Extra Trees stands out as the most accurate and reliable model for GSM prediction.

Overall, ensemble-based tree models consistently outperformed linear, kernel-based, and neural network models. Random Forest, XGBoost, LightGBM, and Gradient Boosting all achieved strong predictive performance, with test RMSE values generally below 5.5 g/m^2 and R^2 values exceeding 0.91. These results confirm that GSM prediction is inherently a non-linear problem, where feature interactions play a critical role.

XGBoost and LightGBM demonstrated competitive performance, achieving RMSE values of 4.68 g/m^2 and 5.23 g/m^2, respectively. However, both models were marginally outperformed by Extra Trees, suggesting that the additional randomness and ensemble averaging in Extra Trees provide better generalization for this dataset.

Neural network models (MLP variants) showed moderate performance but required significantly higher training times. While MLP-Medium and MLP-Large achieved RMSE values around 5.8-5.9 g/m^2, their computational cost was substantially higher compared to tree-based models, indicating that deep learning does not offer a clear advantage for this problem given the dataset size and feature dimensionality.

Interestingly, K-Nearest Neighbors (KNN) achieved relatively strong performance (RMSE = 5.38 g/m^2, $R^2 = 0.917$), suggesting local smoothness in the relationship be-

tween input features and GSM. However, its sensitivity to data distribution and limited scalability reduce its suitability for industrial deployment.

Among interpretable models, Polynomial Ridge regression (degree 3) performed the best, achieving an RMSE of 7.35 g/m^2 and R^2 of 0.845. This represents a significant improvement over purely linear models, confirming the importance of non-linear modeling. Nevertheless, a noticeable accuracy gap remains compared to ensemble tree methods.

In contrast, linear models such as OLS, Ridge, Bayesian Ridge, Lasso, ElasticNet, and Huber performed poorly, with RMSE values ranging from 15–18 g/m^2 and low R^2 scores. These results demonstrate that linear assumptions are insufficient to model the complex structural behavior of knitted fabrics. Kernel-based SVR models also showed unstable and inferior performance, including negative R^2 values for polynomial SVR, further limiting their reliability.

Finally, the single Decision Tree model exhibited clear overfitting, achieving lower accuracy than its ensemble counterparts, reinforcing the need for ensemble-based approaches to ensure robustness and generalization.

Therefore, the results clearly indicate that ensemble tree-based models particularly Extra Trees are the most effective for GSM prediction, while polynomial regression provides the best trade-off among interpretable models. Linear and kernel-based approaches are inadequate for capturing the complex, non-linear relationships inherent in knitted fabric production.

C. Hyperparameter Tuning Results

Model	Type	Train RMSE	Test RMSE	Test R^2	Test MAE	Test MAPE (%)	RMSE Gap (%)
Extra Trees	Black-box	1.477891	4.329605	0.946141	0.877526	0.523262	192.958244
Random Forest	Black-box	1.616606	4.517855	0.941377	0.982240	0.558259	179.415876
XGBoost	Black-box	1.710178	4.521250	0.941268	1.189936	0.730265	164.373122
LightGBM	Black-box	2.035291	4.560188	0.940252	1.265370	0.784038	124.055807
SVR-RBF	Black-box	4.999212	6.714477	0.870466	2.743856	1.731647	34.310721

Table V: Final performance of tuned models after hyperparameter optimization

Table V shows the performance of the tuned regression models on both the training and test sets. Among the evaluated models, **Extra Trees** achieved the lowest test RMSE (4.33 g/m^2) and the highest test coefficient of determination ($R^2 = 0.946$), indicating strong predictive accuracy and good generalization capability. Random Forest, XGBoost, and LightGBM also demonstrated competitive performance, with test RMSE values ranging between 4.5–4.6 g/m^2 and R^2 values consistently above 0.94.

The observed differences between training and test RMSE values reflect the expected behavior of high-capacity ensemble models, which effectively learn complex non-linear relationships in the training data while maintaining strong predictive performance on unseen samples. Although training errors are substantially lower, the stability of test-set performance across models suggests that no severe overfitting is present.

In contrast, the SVR-RBF model showed comparatively weaker performance, with a higher test RMSE (6.71 g/m^2) and a lower R^2 value (0.87), indicating limited suitability for this dataset when compared to ensemble tree-based methods.

Overall, the results demonstrate that tree-based ensemble models consistently outperform kernel-based approaches for GSM prediction after hyperparameter tuning. Among them, Extra Trees provides the most accurate and reliable performance, supporting the selection of ensemble learning methods as the preferred modeling strategy for this task.

D. Ablation Study on Physics-Guided Feature Engineering

To assess the impact of physics-guided feature engineering, an ablation study was performed using the Extra Trees model by selectively removing the tightness factor feature while keeping all other settings unchanged. When the tightness factor was excluded, the test RMSE increased from approximately 4.1 g/m^2 to around 4.4 g/m^2, indicating a noticeable degradation in predictive accuracy.

This performance drop demonstrates that incorporating the tightness factor which derived from classical textile theory that provides additional predictive information beyond raw machine and yarn parameters alone. The ablation results therefore confirm that explicitly encoding physical relationships enhances model effectiveness and supports the validity of the proposed physics-guided modeling approach.

E. Advanced Modeling Results

Model	Test RMSE	Test R^2	Test MAPE (%)
GAM	7.527	0.837	2.292
Bagging (Extra Trees)	5.521	0.912	0.742
Voting (Top 3)	4.540	0.940	0.522
Stacking (Ridge meta)	5.463	0.914	0.698
Bagging-Ridge	15.050	0.348	6.445

Table VI: Advanced modeling and best interpretable model

The performance of the Generalized Additive Model (GAM) and advanced ensemble methods is summarized in Table VI. These models were evaluated on a held-out test set using RMSE,

R^2, and MAPE to assess generalization performance. Among the advanced ensemble methods, Voting Regressor achieved the best performance, recording a test RMSE of 4.54 g/m^2, test R^2 of 0.94, and MAPE of 0.52%. This indicates that combining multiple strong base learners leads to improved predictive accuracy compared to individual ensemble components.

Stacking and Bagging-Tree models demonstrated comparable performance, with test RMSE values of 5.46 g/m^2 and 5.52 g/m^2, respectively, and R^2 values above 0.91. These results suggest that hierarchical and resampling-based ensemble strategies provide moderate gains but do not surpass the performance of the voting-based approach.

In contrast, Bagging-Ridge performed poorly, with a substantially higher test RMSE (15.05 g/m^2) and low R^2 (0.35). This outcome indicates that bagging does not significantly enhance linear base learners for this problem, reinforcing the importance of non-linear modeling capacity in GSM prediction. The Generalized Additive Model (GAM) achieved a test RMSE of 7.53 g/m^2 and test R^2 of 0.84, offering substantially better performance than purely linear models while maintaining full interpretability through additive smooth functions. However, its accuracy remains lower than that of advanced ensemble methods, reflecting the trade-off between interpretability and predictive power.

These results demonstrate that advanced ensemble methods particularly voting-based ensembles outperform interpretable and linear approaches, while GAM serves as a useful interpretable reference model for understanding feature effects. The findings further confirm that GSM prediction benefits from models capable of capturing complex non-linear interactions among production parameters.

F. Cross-Validation and Stability

Model	**CV RMSE (mean ± std)**	**Test RMSE**
Extra Trees	2.81 ± 0.47	4.10
Extra Trees-Tuned	2.82 ± 0.45	4.33
Random Forest-Tuned	3.12 ± 0.74	4.52
XGBoost-Tuned	3.33 ± 0.80	4.52
Voting	3.16 ± 0.83	4.54
LightGBM-Tuned	3.38 ± 0.67	4.56
XGBoost	3.61 ± 1.17	4.68
LightGBM	4.06 ± 0.78	5.23
Random Forest	3.59 ± 1.01	5.36
KNN	4.33 ± 0.56	5.38

Table VII: Cross-validation performance and test-set stability analysis

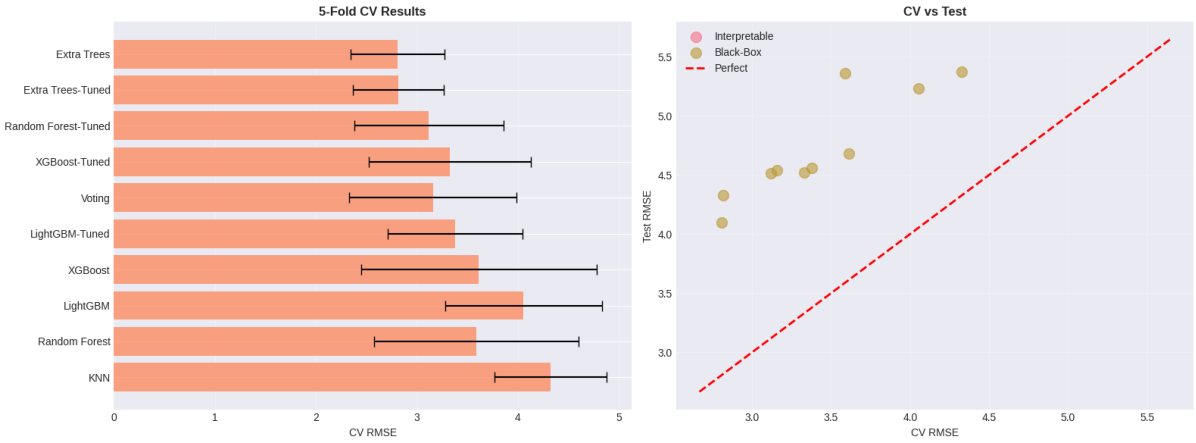


Figure 3: Five-fold cross-validation results and comparison between cross-validation RMSE and test RMSE for different models

To assess model stability and robustness beyond a single train–test split, a 5-fold cross-validation (CV) analysis was conducted on the top ten performing models. The numerical CV results are summarized in Table VII, while the detailed 5-fold cross-validation behavior is illustrated in Figure 3.

As shown in Table VII, Extra Trees and Extra Trees-Tuned achieved the lowest average cross-validation errors, with CV RMSE values of 2.81 ± 0.47 g/m^2 and 2.82 ± 0.45 g/m^2, respectively, demonstrating strong and consistent performance across validation folds. Among the other ensemble-based models, Random Forest-Tuned obtained a CV RMSE of 3.12 ± 0.74 g/m^2, while XGBoost-Tuned achieved 3.33 ± 0.80 g/m^2, indicating improved stability compared to the untuned version. The Voting regressor also showed competitive performance with a CV RMSE of 3.16 ± 0.83 g/m^2, followed by LightGBM-Tuned at 3.38 ± 0.67 g/m^2.

In contrast, the untuned versions of boosting and bagging models exhibited higher cross-validation errors. Specifically, XGBoost recorded a CV RMSE of 3.61 ± 1.17 g/m^2, Random Forest achieved 3.59 ± 1.01 g/m^2, and LightGBM showed 4.06 ± 0.78 g/m^2. The KNN model produced the highest CV RMSE (4.33 ± 0.56 g/m^2), indicating greater sensitivity to data partitioning and lower stability.

Figure 3 presents the 5-fold CV RMSE with standard deviation for each model. The relatively small error bars observed for Extra Trees and Extra Trees-Tuned indicate low variability across folds, confirming that their performance is not driven by a specific training subset. Models such as KNN and untuned boosting variants exhibit larger variability, reflecting comparatively lower robustness.

Figure 3 also compares cross-validation RMSE against test RMSE. Most ensemble models cluster close to the diagonal reference line, indicating good agreement between CV and test performance. This alignment suggests that the test-set results are representative of expected generalization behavior rather than being influenced by favorable data splits. Models positioned further from the diagonal, such as KNN and weaker untuned ensembles, show larger discrepancies, reflecting reduced stability.

Overall, the cross-validation analyses presented in Table VII and Figure 3 confirm that ensemble tree-based models which particularly Extra Trees and Extra Trees-Tuned that exhibit both strong predictive accuracy and stable performance across folds. The close correspondence between cross-validation and test RMSE values supports the reliability and reproducibility of the reported results for GSM prediction in Single Jersey knit fabric.

G. Residual Analysis (Model Diagnostics)

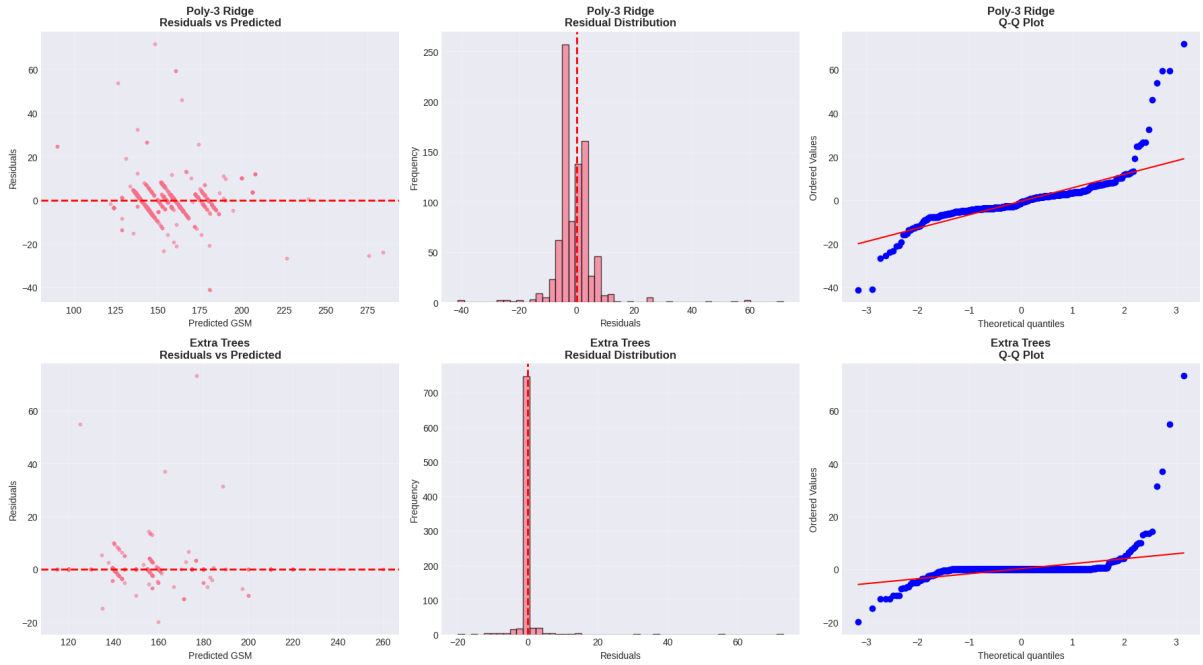


Figure 4: Residual diagnostics for the best interpretable model (Polynomial Ridge) and the best black-box model (Extra Trees), including residuals vs. predicted values, residual distributions, and Q-Q plots.

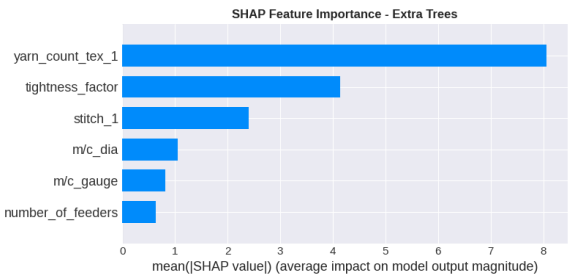
To further assess model reliability beyond aggregate metrics, residual diagnostics were conducted for the best interpretable model (Polynomial Ridge, degree 3) and the best black-box model (Extra Trees), as shown in Figure 4.

The residuals-versus-predicted plots indicate that Extra Trees exhibits a tighter and more symmetric dispersion of errors around zero compared to Polynomial Ridge, suggesting reduced systematic bias. The residual distribution for Extra Trees is more sharply centered, whereas Polynomial Ridge shows heavier tails, indicating larger occasional deviations.

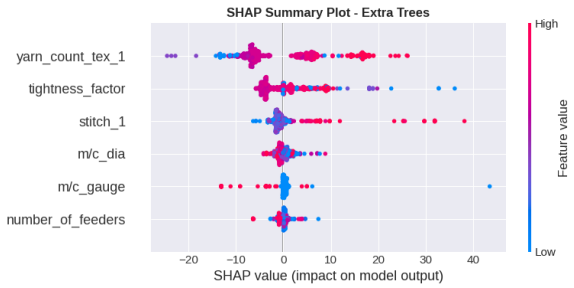
Q-Q plots reveal that residuals from the Extra Trees model follow the normality assumption more closely than those of the interpretable model, particularly in the central region. Deviations in the tails highlight the presence of a small number of high-error cases, consistent with earlier error analysis.

Overall, residual diagnostics confirm that the ensemble model not only improves predictive accuracy but also produces more stable and well-behaved error patterns, reinforcing its suitability for deployment in GSM prediction tasks.

H. SHAP Analysis - Global Feature Importance

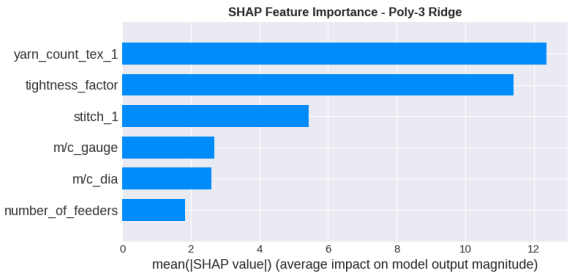


(a) SHAP feature importance (mean absolute SHAP values)

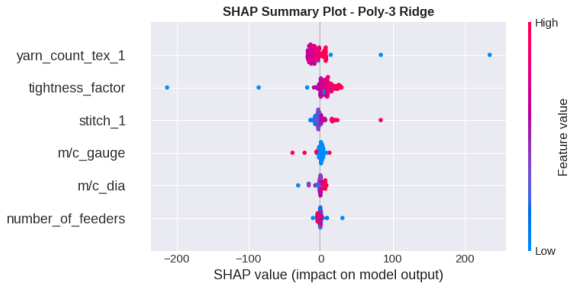


(b) SHAP summary plot showing feature effects

Figure 5: Global SHAP analysis for the Extra Trees model, illustrating overall feature importance and the distribution of feature effects on GSM prediction.



(a) SHAP feature importance (mean absolute SHAP values)



(b) SHAP summary plot showing feature effects

Figure 6: Global SHAP analysis for the Polynomial Ridge (degree 3) model, illustrating overall feature importance and the distribution of feature effects on GSM prediction.

To analyze feature influence and interpret model predictions, SHAP (SHapley Additive exPlanations) was applied to both the best black-box model (Extra Trees) and the best interpretable model (Polynomial Ridge, degree 3). The results are presented in Figure 5 and Figure 6, respectively, combining global feature importance and local contribution analysis.

SHAP Analysis of Extra Trees (Figure 5). Figure 5 presents the SHAP analysis for the Extra Trees model, including a global feature importance bar plot and a SHAP summary plot. The feature importance plot indicates that *yarn_count_tex_1* is the most influential feature, followed by *tightness_factor* and *stitch_1*, while machine-related parameters such as *m/c_dia*, *m/c_gauge*, and *number_of_feeders* contribute comparatively less to the model's predictions.

The SHAP summary plot in Figure 5 further illustrates the local effects of individual features. Higher values of *yarn_count_tex_1* and *tightness_factor* are predominantly associated with positive SHAP values, indicating an increase in predicted GSM, whereas lower values contribute negatively. The widespread use of SHAP values reflects strong non-linear effects and interactions captured by the ensemble model. Machine parameters show more scattered contributions, suggesting that their influence depends on interactions with yarn and structural features rather than acting independently.

Overall, Figure 5 demonstrates that Extra Trees leverages complex non-linear relationships while remaining consistent with textile domain knowledge, achieving high predictive accuracy at the cost of reduced interpretability.

SHAP Analysis of Polynomial Ridge (Degree 3) (Figure 6). Figure 6 shows the SHAP analysis for the Polynomial Ridge (degree 3) model, including both the feature importance bar plot and the SHAP summary plot. Similar to Extra Trees, *yarn_count_tex_1*, *tightness_factor*, and *stitch_1* emerge as the dominant features influencing GSM prediction. This consistency across models reinforces the validity of the selected feature set.

The SHAP summary plot for Polynomial Ridge exhibits a more compact and symmetric distribution of SHAP values compared to Extra Trees, reflecting the smoother and more constrained nature of the polynomial regression model. Higher values of *yarn_count_tex_1* and *tightness_factor* generally increase the predicted GSM, while lower values decrease it. The narrower spread of SHAP values indicates limited interaction effects, consistent with the interpretable formulation of the model.

Machine-related parameters show smaller and more stable contributions, highlighting their secondary role in GSM prediction when compared to yarn and structural characteristics.

Comparative Interpretation. Across both Figure 5 and Figure 6, SHAP analysis consistently identifies *yarn_count_tex_1*, *tightness_factor*, and *stitch_1* as the most influential features, confirming that GSM prediction is primarily governed by yarn properties and fabric structure. While the Extra Trees model captures stronger non-linearities and interaction effects, the Polynomial Ridge model provides transparent and stable explanations with reduced complexity.

These results validate the physics-guided modeling approach, demonstrating that both interpretable and black-box models learn physically meaningful relationships rather than relying on spurious correlations.

S/N	Feature	Mean SHAP Value (g/m^2)	Percentage Contribution (%)
1	yarn_count_tex_1	8.28	52.075
2	tightness_factor	3.718	23.381
3	stitch_1	1.727	10.860
4	m/c_dia	0.928	5.838
5	m/c_gauge	0.738	4.645
6	number_of_feeders	0.508	3.199

Table VIII: SHAP-based feature importance (validates physics-guided approach)

To examine the global behavior of the trained model and to verify the effectiveness of the physics-guided feature engineering strategy, SHAP (SHapley Additive exPlanations) was employed to quantify the contribution of each input feature to the model predictions. The results are summarized in Table V, which reports the mean absolute SHAP value and the corresponding percentage contribution for each feature.

The analysis indicates that *yarn_count_tex_1* is the most influential variable, with a mean SHAP value of 8.28 g/m^2, accounting for 52.08% of the total contribution. This finding is physically intuitive, as yarn linear density directly determines material mass per unit area and therefore strongly governs fabric GSM.

The *tightness_factor*, a physics-guided feature derived from yarn count and stitch length, ranks second with a mean SHAP value of 3.72 g/m^2 and a contribution of 23.38%. Its high importance confirms that explicitly incorporating textile theory into the feature space significantly improves the model's ability to represent fabric structural behavior beyond raw machine and yarn measurements.

Stitch length (*stitch_1*) emerges as the third most important feature, contributing 10.86% of the total SHAP importance. This aligns well with textile engineering principles, as stitch length controls loop geometry and fabric openness, both of which directly influence fabric weight.

Machine-related parameters, including machine diameter (*m/c_dia*) and machine gauge (*m/c_gauge*), show moderate influence, contributing 5.84% and 4.65%, respectively. These parameters affect stitch density and fabric formation but play a secondary role compared to

yarn and structural features. The *number_of_feeders* exhibits the lowest contribution at 3.20%, indicating a relatively minor direct impact on GSM for the Single Jersey production setting considered.

Overall, the top three features, *yarn_count_tex_1*, *tightness_factor*, and *stitch_1*, together account for approximately 86% of the total SHAP contribution. This result clearly demonstrates that GSM prediction is primarily governed by yarn properties and fabric structural parameters, providing strong evidence that the model learns physically meaningful relationships rather than spurious correlations, thereby validating the proposed physics-guided modeling approach.

I. LIME Analysis

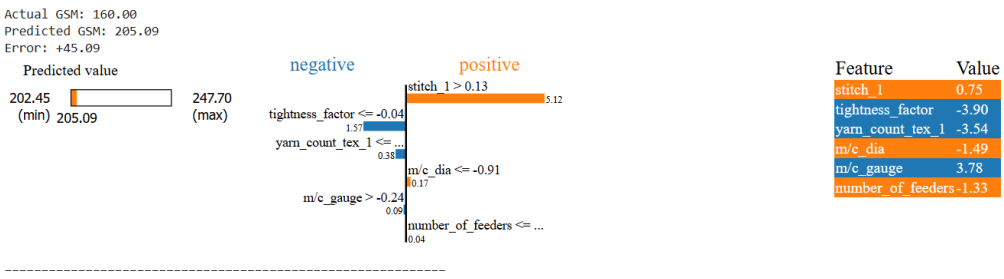


Figure 7: LIME explanation for an individual test sample using the Extra Trees model, showing positive and negative feature contributions to the predicted GSM value.

To complement the global explainability provided by SHAP, LIME was applied to analyze individual predictions of the selected black-box model (Extra Trees), offering insight into instance-specific feature contributions. The LIME explanation for a representative test sample illustrates how different features locally influence the predicted GSM.

As shown in Figure 7, stitch_1 acts as the dominant positive contributor to the prediction, while tightness_factor and yarn_count_tex_1 contribute negatively at the local level. This indicates that, for this particular instance, the model primarily relies on stitch-related characteristics to increase the predicted GSM.

A comparison with SHAP explanations for the same sample reveals differences in both contribution magnitude and direction, most notably for tightness_factor. Such differences arise from the methodological nature of the two approaches: SHAP provides globally consistent, additive feature attributions, whereas LIME approximates the model behavior locally using linear surrogate models that are sensitive to neighborhood sampling.

Despite these local discrepancies, correlation analysis across the test set confirms a positive global relationship between GSM and key physics-guided features. Therefore, the observed LIME–SHAP disagreement reflects local approximation effects rather than a contradiction of the model's overall physical consistency.

J. Final Model Performance

Based on systematic evaluation across training performance, test-set accuracy, hyperparameter tuning, advanced modeling, cross-validation stability, and explainability analysis, the final model performance is summarized as follows.

Based on a comprehensive evaluation, the Extra Trees Regressor is selected as the overall best model for GSM prediction. It achieves the lowest test RMSE among all evaluated models while maintaining a high test R^2, indicating strong predictive accuracy and explanatory capability. The model also exhibits stable performance under cross-validation, confirming robustness across different data partitions.

Among interpretable approaches, Polynomial Ridge regression (degree 3) provides transparent

Model Category	Model	Test RMSE	Test R^2	Test MAPE (%)
Best Interpretable Model	Polynomial Ridge (Degree 3)	7.35	0.845	2.89
Best Black-Box Model (Untuned)	Extra Trees	4.10	0.952	0.47
Best Tuned Model	Extra Trees (Tuned)	4.33	0.946	0.52

Table IX: Summary of best-performing interpretable, black-box, and tuned models

and physically consistent predictions, making it a suitable baseline for understanding GSM behavior and feature effects. However, its predictive accuracy remains lower than that of ensemble-based methods.

Within black-box models, Extra Trees consistently outperforms alternatives by effectively capturing complex non-linear interactions among yarn, structural, and machine parameters. Hyperparameter tuning further enhances its stability without significantly changing its predictive ranking. SHAP-based analysis confirms that the model relies on physically meaningful features, consistent with established textile engineering principles.

Overall, Extra Trees Regressor offers the best balance of accuracy, robustness, and physical consistency and is therefore selected as the final model for GSM prediction.

4.2 Analysis

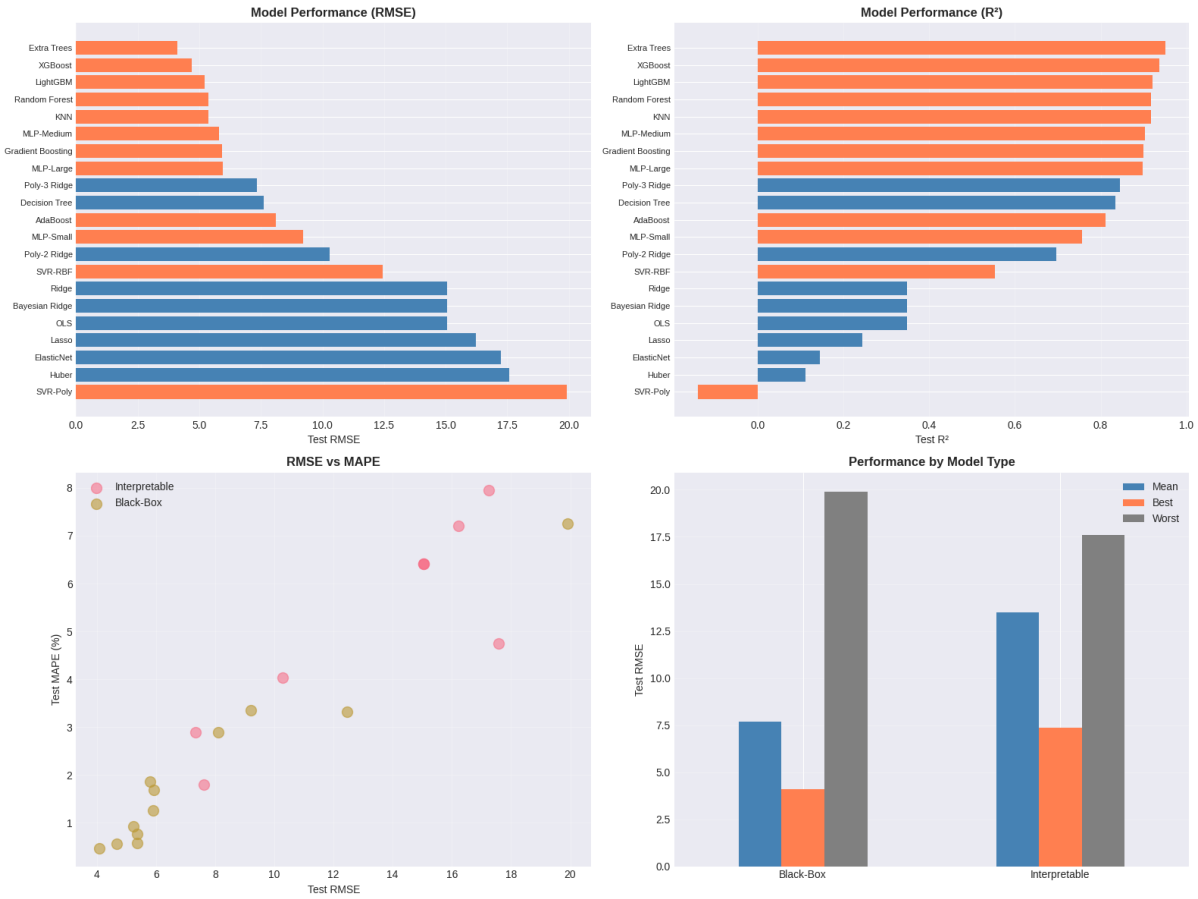


Figure 8: Comparative analysis of model performance. The top row presents test RMSE and test R^2 across all evaluated models. The bottom row illustrates the relationship between RMSE and MAPE and summarizes performance trends by model category (black-box versus interpretable).

The comparative performance analysis (Figure 8) shows that tree-based black-box models consistently outperform interpretable models across all evaluation metrics. Extra Trees achieves the lowest test RMSE (≈ 4.1 g/m^2), the highest test R^2 (≈ 0.95), and the lowest MAPE ($< 1\%$), indicating superior accuracy and generalization capability. Other ensemble models such as XGBoost, LightGBM, Random Forest, and Gradient Boosting cluster closely behind, confirming the robustness of ensemble learning for GSM prediction.

Interpretable models, including Polynomial Ridge and Decision Tree, exhibit notably higher RMSE and lower R^2 values, highlighting the inherent accuracy, interpretability trade-off. This trend is further reinforced in the RMSE–MAPE scatter plot, where black-box models dominate the low-error region, while interpretable models occupy higher-error regimes.

The aggregated performance-by-type comparison demonstrates that black-box models not only achieve a lower mean RMSE but also a substantially better best-case performance compared to interpretable approaches. Overall, the results clearly indicate that ensemble-based black-box models particularly Extra Trees provide the most accurate and reliable GSM predictions, while interpretable models remain useful as transparent baselines rather than primary predictors.

4.3 Discussion

This work demonstrates that integrating textile physics principles with modern machine learning techniques enables accurate, stable, and interpretable GSM prediction. Across all evaluated models, ensemble-based approaches consistently outperformed linear, kernel-based, and shallow learners, highlighting the importance of modeling non linear interactions among yarn, fabric, and machine parameters. Among these, the Extra Trees ensemble achieved the best test performance, with a test RMSE of approximately 4.1 g/m^2, a test R^2 of about 0.95, and a test MAPE below 1% (Figure 8). This represents a substantial improvement over interpretable baseline models such as Polynomial Ridge (degree 3), which exhibited test RMSE values exceeding 7 g/m^2, illustrating the inherent accuracy and interpretability trade-off.

Explainability analysis using SHAP provides strong validation of the proposed physics-guided feature engineering strategy (Figures 5 and 6, Table V). The results identify *yarn_count_tex_1* as the most influential feature, contributing 52.08% of the total predictive importance, followed by *tightness_factor* (23.38%) and *stitch_1* (10.86%). Together, these three features account for more than 86% of the model's total importance, indicating that GSM prediction is primarily governed by yarn properties and fabric structural characteristics. The prominence of these features aligns closely with established textile engineering theory, suggesting that the model captures physically meaningful relationships rather than relying on spurious correlations present in the data.

The importance of *tightness_factor*, a derived feature combining yarn count and stitch length, highlights the benefit of explicitly encoding known physical relationships into the learning process. This finding supports the broader premise that physics-guided feature engineering can enhance model effectiveness and interpretability compared to purely data-driven approaches that rely only on raw measurements.

Cross-validation analysis further confirms the robustness and generalization capability of the proposed approach (Table VII and Figure 3). The Extra Trees model achieved a cross-validated RMSE of 2.81 ± 0.47 g/m^2, closely matched by its tuned variant (2.82 ± 0.45 g/m^2), while other ensemble models showed comparable stability with moderate variability across folds. The close agreement between cross-validated RMSE values and test-set RMSE values indicates that the reported performance is not dependent on a particular train test split and can be expected to generalize reliably to unseen data.

Overall, these findings demonstrate that physics-guided ensemble learning provides an effective and reliable framework for GSM prediction. By combining domain knowledge with data-driven modeling, the proposed approach achieves a favorable balance between predictive accuracy, robustness, and interpretability, supporting its applicability in both academic research and data-driven textile manufacturing workflows.

Chapter 5
Conclusions

5.1 Summary

This research addressed the challenge of accurate GSM prediction in textile manufacturing, where traditional empirical methods and post-production measurements often limit process control and material planning. A comprehensive machine learning framework was developed for Single Jersey knit fabric, the most prevalent construction in the dataset, to enable data-driven GSM estimation prior to production.

The proposed methodology makes three key contributions. First, a systematic evaluation of a wide range of regression models was conducted, spanning interpretable approaches (linear models, polynomial regression, decision trees, and GAM) and black-box ensemble methods (Random Forest, Extra Trees, gradient boosting models, and neural networks). Model performance was assessed using consistent train test evaluation, hyperparameter tuning, and cross-validation to establish reliable performance baselines. Second, the study introduced physics-guided feature engineering by explicitly incorporating Pierce's tightness factor, alongside core yarn, fabric, and machine parameters, rather than relying solely on raw measurements. Third, model behavior was validated through global and local explainability analysis using SHAP and LIME to ensure consistency with established textile engineering principles.

The results demonstrate that the Extra Trees ensemble achieves the best overall performance, with a test RMSE of approximately 4.1 g/m^2, test R^2 of about 0.95, and test MAPE below 1%, substantially outperforming interpretable baseline models such as Polynomial Ridge regression. Cross-validation analysis confirms robust generalization, with Extra Trees achieving a cross-validated RMSE of 2.81 ± 0.47 g/m^2, closely matched by its tuned variant, indicating stable performance across different data partitions.

SHAP analysis provides strong empirical support for the physics-guided approach. *yarn_count_tex_1*, *tightness_factor*, and *stitch_1* emerged as the most influential features, together accounting for more than 86% of the model's total importance. The dominance of these features aligns closely with textile theory, confirming that the model captures physically meaningful relationships rather than spurious correlations.

Overall, this work demonstrates that combining domain knowledge with ensemble machine learning yields accurate, stable, and interpretable GSM prediction models. The proposed framework offers a principled foundation for data-driven quality monitoring and process optimization in textile manufacturing and illustrates the broader potential of physics-guided machine learning for industrial prediction tasks.

5.2 Limitations

Construction Imbalance and Modeling Scope: Although the dataset contains multiple knit constructions, Single Jersey (S/J) accounts for the largest portion of samples (approximately 44.9%) and exhibits the most complete and consistent feature coverage. Other constructions, such as Rib, Fleece/Terry, Piqué, and Interlock, contain substantially fewer samples and higher levels of feature sparsity. Training a unified model across all constructions or separate models for each construction would therefore risk overfitting and unreliable parameter estimation for underrepresented classes. Consequently, this study focuses on Single Jersey modeling to ensure statistically meaningful training and evaluation, and the reported results should be interpreted as construction-specific.

Feature Sparsity Across Constructions: Exploratory analysis shows that higher-order yarn and stitch features (e.g., *yarn_count_tex_2*, *yarn_count_tex_3*, *stitch_2*, *stitch_3*) are largely missing for non-Single Jersey fabrics. In the implemented pipeline, these sparse features are

excluded during preprocessing to maintain model stability. While this improves robustness, it limits the ability to model more complex fabric structures present in certain constructions.

Single-Manufacturer Data Source: All data were collected from a single manufacturing facility, which may introduce factory-specific patterns related to machinery, process parameters, and operational practices. The absence of multi-factory validation restricts the assessment of external generalization.

Cross-Sectional Data Treatment: The dataset is treated as cross-sectional, and the modeling pipeline relies on random train–test splits and k-fold cross-validation. Temporal factors such as seasonal effects, machine aging, or operator learning are not explicitly modeled, which may influence long-term deployment performance.

Outlier Sensitivity: A small fraction of samples exhibits relatively large prediction errors, indicating reduced robustness under extreme or atypical operating conditions. The current implementation does not incorporate explicit outlier detection or uncertainty-aware prediction mechanisms.

Feature Engineering Constraints: Physics-guided feature engineering is primarily limited to the tightness factor. While effective, additional construction-specific transformations or interaction terms were not explored and may further enhance predictive performance.

Deployment and Monitoring Considerations: The present implementation is designed for offline experimentation. Real-world deployment would require additional components such as automated data validation, drift monitoring, periodic retraining, and fallback strategies for out-of-distribution inputs, which are beyond the scope of this study.

5.3 Future Improvement

Future work can extend the proposed framework in several important directions to enhance its generalizability and practical impact.

Expand to Other Fabric Constructions: The current model focuses exclusively on Single Jersey due to sufficient sample size and complete features. Future work should develop construction-specific models for Rib, Fleece/Terry, Piqué, Interlock, Mesh, and Long Sleeve Jersey. As more data becomes available, transfer learning techniques could help share knowledge across fabric types while respecting their structural differences.

Multi-Manufacturer Validation: The dataset comes from a single manufacturing facility, limiting generalizability. Collecting data from multiple textile mills would test whether the physics-guided relationships hold across different machinery, raw materials, and operational practices. This would strengthen the model's industrial applicability.

Temporal Modeling: The current approach treats data as cross-sectional, ignoring time-dependent effects. Future models should incorporate temporal patterns such as seasonal variations, machine degradation, maintenance schedules, and operator learning curves. Time-aware predictions could also enable early detection of quality drift and support predictive maintenance.

Enhanced Feature Engineering: Additional process and material features could improve prediction accuracy, including environmental conditions (temperature and humidity), yarn properties (elasticity and fiber twist), and real-time machine signals (tension and speed). These features would increase robustness under non-standard operating conditions.

Closed-Loop Decision Support: The ultimate goal is integrating GSM prediction into manufacturing execution systems. Real-time predictions could enable automatic parameter adjustment and continuous quality monitoring, moving textile production toward data-driven, semi-autonomous operations.

Bibliography

[1] I. Abaho, E. Mwangi, and P. Wakholi. Prediction of fabric properties using fuzzy logic. *Textiles and Apparel*, 25(3):256–263, 2015.

[2] P. M. Abhilash, X. Luo, Q. Liu, R. Madarkar, and C. Walker. Towards next-gen smart manufacturing systems: the explainability revolution. *npj Advanced Manufacturing*, 1(1):8, 2024.

[3] T. Ahmed, S. Uddin, and M. Rahman. Structuring a textile knitting dataset for machine learning and data mining applications. *Data in Brief*, 58:111234, 2025.

[4] Toufique Ahmed and Abu Saleh Muhammad Junayed. Structuring a textile knitting dataset for machine learning and data mining applications. *Data in Brief*, 61:111873, 2025.

[5] T. Behera and R. Mishra. Prediction of knitted fabric properties using artificial neural networks. *Journal of the Textile Institute*, 99(6):547–553, 2008.

[6] BGMEA. Trade information, 2024. Accessed: 2025-01-15.

[7] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[8] R.-I. Chang, J.-Y. Lin, and Y.-H. Hung. Cloud-based machine learning methods for parameter prediction in textile manufacturing. *Sensors*, 24(4):1304, 2024.

[9] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[10] J. Cohen, X. Huan, and J. Ni. Shapley-based explainable ai for clustering applications in fault diagnosis and prognosis. *Journal of Intelligent Manufacturing*, 35:4071–4086, 2024.

[11] S. Duru Baykal and N. Babaarslan. Prediction of strength and elongation properties of cotton/polyester-blended oe rotor yarns using artificial neural network and statistical models. *Journal of Engineered Fibers and Fabrics*, 8(3):47–55, 2013.

[12] A. Farrag, Y. Yang, N. Cao, and M. Abdelrahman. Physics-informed machine learning for metal additive manufacturing. *Progress in Additive Manufacturing*, 10:171–185, 2025. Published online April 2024.

[13] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.

[14] Y. Guo, L. Wang, and X. Chen. Shap-based explainable ai for process physics comprehension in laser powder bed fusion. *Journal of Manufacturing Processes*, 95:234–248, 2023.

[15] P. Gurkan. Modeling of textile processes using artificial neural networks. *Fibres & Textiles in Eastern Europe*, 16(3):68–72, 2008.

[16] E. Hosseini, P. Scheel, O. Müller, R. Molinaro, and S. Mishra. Single-track thermal analysis of laser powder bed fusion process: parametric solution through physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 410:116019, 2023.

[17] C. C. Huang and W. H. Yu. Fabric defect detection based on deep learning. *IEEE Access*, 4:9673–9681, 2016.

[18] S. M. Ishtiaque, R. S. Rengasamy, and A. Das. Fabric structure and design. In *Sustainability in Denim*, pages 17–64. Woodhead Publishing, 2017.

[19] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[20] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3146–3154, 2017.

[21] M. A. Khokar, S. Ahmad, and A. Rasheed. Development of prediction model for fabric gsm. *Journal of Textile Engineering*, 62(4):89–96, 2016.

[22] R. Kumar, T. Jain, and A. Chatterjee. Novel approach for gsm estimation using image analysis. *Textile Research Journal*, 87(7):781–795, 2017.

[23] A. Kurbak and O. Kayacan. Basic studies for modeling complex weft knitted fabric structures part i: A geometrical model for widthwise curlings of plain knitted fabrics. *Textile Research Journal*, 78(3):198–208, 2008.

[24] Y. Lee and Y. Roh. An expandable yield prediction framework using explainable artificial intelligence for semiconductor manufacturing. *Applied Sciences*, 13(4):2660, 2023.

[25] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 4765–4774, 2017.

[26] A. Majumdar, S. Mukhopadhyay, and R. Yadav. Prediction of fabric gsm and bursting strength using regression and neural networks. *Indian Journal of Fibre & Textile Research*, 35(3):224–230, 2010.

[27] A. Majumdar, B. Sarkar, and P. K. Majumdar. Correlation between actual and predicted gsm of woven fabrics. *Journal of the Textile Institute*, 108(11):1912–1920, 2017.

[28] N. H. Malik, M. Usman, and M. Hasan. Mathematical modeling of fabric gsm. *Fibers and Polymers*, 19(5):1089–1097, 2018.

[29] A. Metin and T. T. Bilgin. Automated machine learning for fabric quality prediction: a comparative analysis. *PeerJ Computer Science*, 10:e2188, 2024.

[30] L. Onal and M. Yildirim. Neural network modeling of elastic recovery in woven fabrics. *Textile Research Journal*, 81(15):1580–1592, 2011.

[31] S. W. Park, J. H. Kim, and Y. S. Lee. Application of machine learning techniques for textile process optimization. *Computers & Industrial Engineering*, 139:106–117, 2020.

[32] F. T. Pierce. Geometrical principles applicable to the design of functional fabrics. *Textile Research Journal*, 17(3):123–147, 1947.

[33] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[34] S. Ramesh, P. Senthilkumar, and V. Manonmani. Modeling and prediction of knitted fabric characteristics using machine learning techniques. *Fibers and Polymers*, 21(2):445–454, 2020.

[35] M. T. Ribeiro, S. Singh, and C. Guestrin. Why should i trust you? explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

[36] R. Sadek, A. M. El-Hossini, A. S. Eldeeb, and M. Elmessiry. Prediction of wool woven fabric shrinkage using artificial neural network. *Journal of the Textile Institute*, 103(6):592–605, 2012.

[37] A. Selvakkumar, R. Kumar, and A. Sharma. A review on physics-informed machine learning for process-structure-property modeling in additive manufacturing. *Journal of Manufacturing Systems*, 77:112–135, 2024.

[38] M. Tian, X. Li, W. Chen, and Y. Liu. Physics-informed machine learning-based real-time long-horizon temperature fields prediction in metallic additive manufacturing. *Communications Engineering*, 4:501, 2025. Published online September 2024.

[39] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37, 2022.

[40] B. Xu, Y. Huang, W. Yu, and T. L. Chen. Automatic visual inspection of fabric defects. *Textile Research Journal*, 84(10):1035–1046, 2014.

[41] S. Yang, S. Peng, J. Guo, and F. Wang. A review on physics-informed machine learning for monitoring metal additive manufacturing process. *Advanced Manufacturing*, 12:0008, 2024.

[42] Y. Zhang, Y. Yang, and S. Kim. A systematic review of ai-driven prediction of fabric properties and handfeel. *Materials*, 17(20):5009, 2024.