STM JOURNALS

Journal of
Polymer & Composites

ISSN: 2321-2810 (Online)
ISSN: 2321-8525 (Print)
Volume 14, Special Issue 3, 2026
DOI (Journal): 10.37591/JoPC

https://journals.stmjournals.com/jopc

Research JoPC

Machine Learning-Based Quantification of Polymer Structure Property Relationships for Predictive Material Design

Kalpana G. Joshi[1], Shanthi Kumaraguru[2], Prashant M. Yawalkar[3], P. William[4,*], Prashant V. Thokal[5], Jaikumar M. Patil[6]

Abstract

Polymer structures exhibit complex, hierarchical arrangements that strongly influence macroscopic properties, yet consistent quantification remains challenging due to nonlinear interactions and limited unified modeling strategies. Existing approaches inadequately capture generalized structure–property mappings across diverse polymer systems. This research aims to establish a machine learning-based quantification model for polymer structure–property relationships to support predictive material design. A Polymer Structure Property Dataset of 5,000 polymer samples includes structural descriptors and experimentally measured thermal and physical properties. Data preprocessing involves Z-score normalization and missing value imputation using K-Nearest Neighbors (KNN). Feature extraction employs Principal Component Analysis (PCA), enhancing discriminative structural patterns while reducing dimensional redundancy. The model systematically collects polymer data, preprocesses and refines the data, it also extracts salient structural features, and feeds them into an optimized predictive engine. The proposed Electric Fish Optimizer-driven Intelligent Support Vector Machine (EFO-ISVM) combines global optimization and adaptive learning. The EFO is utilized for hyperparameter tuning due to its efficient exploration–exploitation balance inspired by electric field sensing behavior, while the ISVM performs nonlinear regression and classification to model complex structure–property dependencies with improved generalization. Quantification is achieved by mapping structural descriptors to target properties using Python, enabling precise evaluation of interdependencies and sensitivity patterns, with an achieved prediction Stability of (97.92%). The model demonstrates enhanced predictive stability, reduced error tendencies, and improved generalization across multiple property indicators. The approach supports reliable material design decisions, offering a scalable pathway for intelligent polymer engineering and advanced composite development.

***Author for Correspondence**
P. William

[1]Assistant Professor, School of Engineering and Technology, Sanjivani University, Kopargaon, Maharashtra, India
[2]Assistant Professor, Department of Information Technology, D Y Patil College of Engineering, Akurdi, Pune, Maharashtra, India
[3]Associate Professor, Department of Computer Engineering, MET's Institute of Engineering, BKC, Nashik, Maharashtra, India
[4]Professor (Research), School of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India
[4]Research Fellow, Centre for Research and Consultancy, UNITAR International University, Selangor, Malaysia
[5]Assistant Professor, Department of Electrical Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India
[6]Associate Professor, Department of Computer Science and Engineering, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, SGBAU, Amravati, Maharashtra, India

Received Date: May 22, 2026
Accepted Date: June 02, 2026
Published Date: June 10, 2026

Citation: Kalpana G. Joshi, Shanthi Kumaraguru, Prashant M. Yawalkar, P. William, Prashant V. Thokal, Jaikumar M. Patil. Machine Learning-Based Quantification of Polymer Structure Property Relationships for Predictive Material Design. Journal of Polymer & Composites. 2026; 14(Special Issue 3): S737–S754p.

Keywords: Polymer structure–property relationships, predictive material design, polymer informatics, molecular feature engineering, property prediction modeling.

INTRODUCTION

The structural property relationships for Predictive Material Design to be characterized as the structure of the polymer, which defines the unique properties of the polymer [1]. For

polymers, characteristics such as length, branch, crystallinity, intermolecular interactions, and chemistry make up the basis for determining their physical properties such as tensile strength, elasticity, thermal stability, electrical properties, and durability. The principle is important in material science due to the ability of explaining why various polymers exhibit varied properties under identical conditions [2].

This field has been extensively applied to advanced materials research in various industries. For healthcare, it provides the means for designing biocompatible and biodegradable materials that utilized as implants or carriers for drugs [3]. In terms of packaging and construction, it facilitates the manufacturing of robust but lightweight materials with high efficiency. Predictive material design, grounded on these correlations, becomes highly significant in advancing innovation and minimizing experimentation [4].

However, some challenges arise from this field. This makes it hard to create a direct and universal relationship between the structure and properties of such systems [5]. It is also challenging to test materials experimentally since it is usually costly and time-consuming, and is subjected to certain restrictions. Also, it should be noted that it is hard to develop materials since the requirements for their properties are conflicting [6].

The benefits of polymer structure-property relationships can be listed as optimized materials, efficient designs of polymers depending on their applications, shorter development times, and greater insights into material behaviors [7]. The drawbacks of this technique consist of extremely complicated analyses, its reliance on large amounts of experimental data, and modeling difficulties regarding realistic polymer properties. This area is of vital importance in terms of developing new materials, although it still faces major challenges in gaining fully reliable results [8].

Problem statement

The description of relationship between polymer structure and property is difficult due to non-linearity, multi-dimensionality, and lack of predictive models capable of describing relationships for different polymers. To address these challenges, this research designs an Electric Fish Optimizer-driven Intelligent Support Vector Machine (EFO-ISVM) model for quantitatively estimating structure-property relationships of polymers, with accurate prediction ability and enhanced generalization capability. It ensures accurate predictions of the properties of polymers, efficient polymer designs, saves on experimental costs, and promotes polymer composites research.

Research Organization

Introduction to polymer structure-property relationships and previous research-based predictive modeling is covered in Sections 1 and 2. The suggested EFO-ISVM approach and its design are described in Sections 3. Experimental results with discussion and conclusion of this research are provided in Sections 4 and 5, and final conclusion in Section 6.

RELATED WORK

The current literature shows that Machine Learning (ML) techniques are effective in improving the structure-property relationship of polymer materials using state-of-the-art methods like Artificial Neural Networks (ANNs), Deep Learning (DL), and optimization algorithms. Table 1 shows the effectiveness of ML techniques.

Recent studies have highlighted the importance of advanced material characterization, sustainable composite development, and data-driven modeling for improving material performance and accelerating polymer design. Reinforcement strategies, interfacial engineering, and machine learning-assisted optimization have demonstrated significant potential for enhancing structure–property relationships in polymer and composite materials [18–20].

Table 1. ML approaches for polymer structure–property prediction.

Ref.	Objective	Method	Key results	Limitations
[9]	Improvement of polymer property prediction by incorporating molecular interaction effects.	ML integrated with Group Interaction Modeling (GIM) to capture molecular-level interactions in polymers.	Higher prediction accuracy achieved by modeling.	Requires high-quality structured datasets.
[10]	Enhancement of physical property prediction for textile polymer composites.	Multi-Objective Optimization (MOO) based ANNs used for nonlinear property learning.	Improved accuracy and stability in predicting composite material physical properties.	High computational cost and strong dependence on experimental data quality.
[11]	Analysis of progress, challenges, and future scope of ML in polymer science.	ML applications in polymer research.	Identified key advancements in ML-based polymer property prediction and material discovery.	Lack of standardized datasets and poor interpretability of complex models.
[12]	Evaluation of ML efficiency in fiber-reinforced polymer applications in structural engineering.	ML techniques applied to Fiber-Reinforced Polymer (FRP) structural analysis.	Improved efficiency in predicting structural performance of FRP materials.	Limited real-world validation and variability in material behavior affects reliability.
[13]	Development of reverse design capability for polymers with desired properties.	Polymer Targeted Optimization Algorithm (PolyTAO) model for inverse polymer design.	Enabled generation of polymer structures based on desired property requirements.	High computational cost and limited scalability for complex polymer systems.
[14]	Prediction of composite material properties using experimental datasets.	ML models trained on experimental composite material data.	Improved prediction accuracy for mechanical and physical properties of composites.	Sensitive to noise and inconsistencies in experimental datasets.
[15]	Prediction of natural fiber polymer composite properties.	DL models applied for property prediction of Natural Fiber Polymer Composites (NFPCs).	High prediction accuracy for natural fiber-based composites.	Requires large datasets and suffers from low interpretability.
[16]	Prediction of polymer physical properties using structure–property relations.	Quantitative Structure–Property Relationship (QSPR) + ML on ATHAS dataset.	Improved prediction of polymer thermal and physical properties.	Limited to ATHAS dataset; depends on descriptor quality.
[17]	Prediction of polymer physical characteristics using data-driven models.	ML regression models on polymer datasets.	Better accuracy than traditional methods.	Dataset dependency and limited generalization.

The growing availability of material datasets and intelligent predictive algorithms has enabled more efficient exploration of complex structure–property relationships, supporting the development of high-performance polymer systems and next-generation composite materials [21–23].

Research Gap

Limitations of existing approaches include lack of generalizability, high computational cost, and dependence on large, quality databases for predicting polymer properties. Moreover, issues related to poor interpretability and scalability hinder current design paradigms [9–15]. In this research, EFO-ISVM technique is more effective in predicting the properties of polymers because of the integration of adaptive hyperparameter optimization and intelligent non-linear learning that leads to increased generalization capability, less prediction error, stability, and accurate modeling of structure-property relations.

Compared with conventional statistical and physics-based modeling approaches, the proposed EFO-ISVM framework offers enhanced capability for capturing nonlinear and high-dimensional structure–property relationships inherent in polymer systems. Statistical models are often limited by linear assumptions, whereas physics-based simulations require extensive computational resources and

domain-specific formulations. By integrating adaptive hyperparameter optimization with intelligent nonlinear learning, EFO-ISVM provides improved prediction accuracy, generalization capability, and computational efficiency for diverse polymer materials.

PROPOSED METHODOLOGY

This section presents the development of a ML model for assessing structure-property relations of polymers in material design. Figure 1 presents the model formulated in this research. The Polymer Structure–Property Dataset is preprocessed involving the K-nearest neighbors (KNN) imputation approach along with Z-score normalization, while feature selection has been done through the application of PCA. The novel aspect of this research is the application of EFO-ISVM for predicting nonlinear relations.

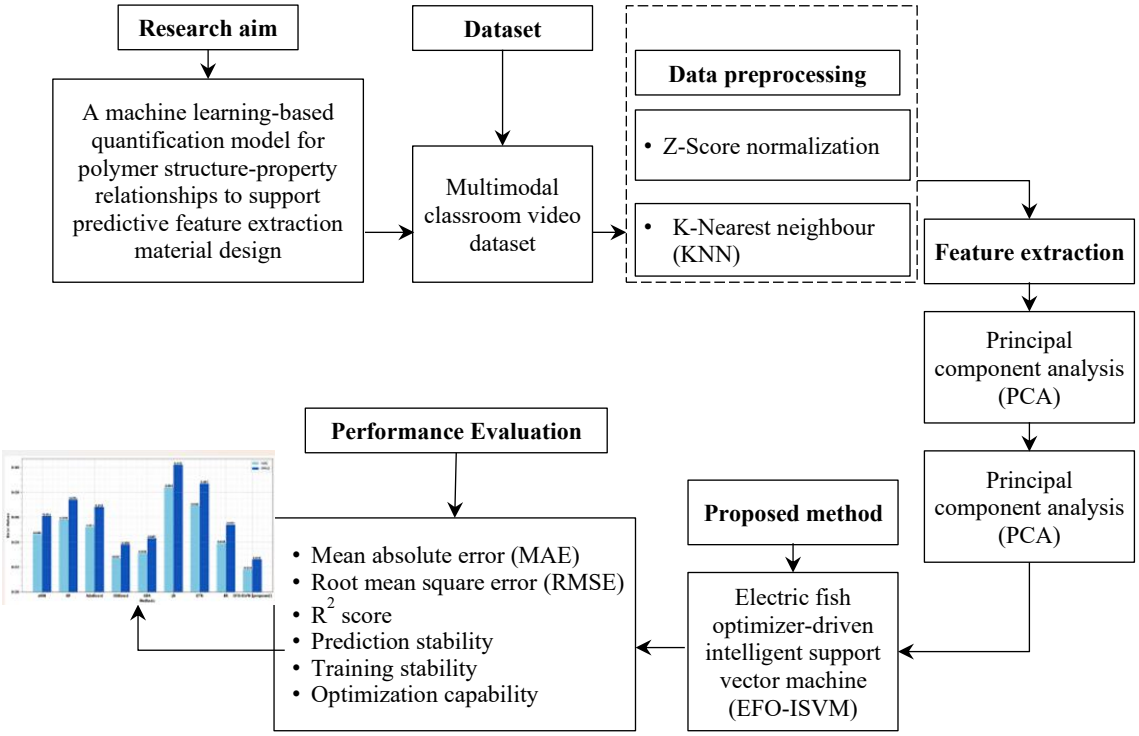


Figure 1. Methodological workflow for predictive polymer material design.

Data Acquisition

The Polymer Structure Property Dataset gathers for this research from Kaggle platform, Link: https://www.kaggle.com/datasets/zara2099/polymer-structure-property-dataset/data. It contains polymer structure and material property information designed for predictive material design and polymer informatics research. The dataset includes 5,000 rows and 28 columns covering polymer structural descriptors, processing conditions, thermal characteristics, mechanical indicators, and material behavior classes. The data represents multiple polymer categories, monomer compositions, crystallinity levels, chain structures, and environmental exposure conditions associated with material performance evaluations. Features: Polymer_ID, Polymer_Type, Monomer_Type, Chain_Structure, Crystallinity_Level, Molecular_Weight, Degree_of_Polymerization, Crosslink_Density, Glass_Transition_Temp_C, Melting_Temperature_C, Density_g_cm3, Processing_Method, Processing_Temperature_C, Cooling_Rate_C_min, Thermal_Conductivity_W_mK, Elastic_Modulus_GPa, Tensile_Strength_MPa, Water_Absorption_Percent, Humidity_Exposure_Percent, UV_Resistance_Index Structural_Complexity_Index, Composite_Filler_Presence, Nano_Reinforcement, Thermal_Stability_Class, Mechanical_Strength_Class, Chemical_Resistance_Class , Electrical_Behavior_Class.

Data Preprocessing

Raw polymer datasets regularly contain incomplete values, inconsistent scales, redundant information, and experimental noise. Preprocessing plays an important role in improving the quality of the data for polymer analysis by employing the KNN method for filling in missing values and Z-score normalization for feature scaling.

Z-Score Normalization for Polymer Structure Descriptor Standardization

Z-score normalization assists in making polymer descriptor variables standardized such that the scale difference which causes the stabilization are eliminated when applying the models for prediction. It is a process whereby the data is normalized by first deducting the mean value of the variable and then dividing the new value by the standard deviation.

$$y_{zscore} = \frac{y_{inst} - u}{\delta} \quad (1)$$

In Equation (1), y_{zscore} denotes the normalized value, y_{inst} represents the original input, u is the mean, and δ is the standard deviation. Therefore, normalized polymer features achieve balanced distributions, reduced variance influence, improved convergence, and enhanced prediction consistency across diverse material properties.

K-Nearest Neighbor (KNN) Imputation for Missing Polymer Data Completion

KNN imputation completes missing polymer descriptor values using neighboring similarities, improving dataset consistency, reliability, and ML prediction accuracy. The KNN method infers missing values of descriptors for polymers through proximity to samples that are alike in terms of distance measures. The KNN preprocessing method in this research, contributes to dataset completion, maintains data structure, minimizes data loss, and increases prediction accuracy and reliability. The Euclidean distance is determined using the following Equation (2).

$$C_{w,z} = \sqrt{weight + squared\ distance\ from\ present\ coordinates} \quad (2)$$

$C_{w,z}$ represents movement cost, where $w$denotes weight factor and z indicates squared distance from present coordinates.

$$weight = \frac{total\ number\ of\ coordinates}{number\ of\ present\ coordinate} \quad (3)$$

In the Equation (3), $weight$ represents the calculated importance or completeness factor of the coordinate data, $total\ number\ of\ coordinates$ denotes the overall expected coordinate positions in the dataset, and $number\ of\ present\ coordinate$ indicates the actual available or detected coordinate points used during computation. Therefore, KNN imputes missing values using similarity-based averaging of nearest data points, preserving local structure and improving dataset completeness.

Principal Component Analysis (PCA) for Polymer Structural Future Extraction

After completed the data preprocessing step, the preprocessed data of the polymer structures proceeds into the feature extraction. In this process, crucial structural features are chosen from the polymer data to develop intricate structure-property correlations.

For reducing dimensionality, removing redundancy among descriptors, and retaining maximum variance to increase efficiency and improve the ability to discriminate. Dimensionality Reduction through PCA is the procedure of reducing dimensionality from the dataset of polymer descriptors to a lower dimensional space. The direction of maximum variability from the dataset is found using the eigenvectors of the covariance matrix. It simplifies the computation while ensuring all important information is preserved.

$$\Psi = \frac{1}{N}\sum_{j=1}^{N} \Gamma_j \quad (4)$$

$$\Phi_j = \Gamma_j - \Psi \tag{5}$$

$$CV_j = v_j V_j \tag{6}$$

Equations (4–6) depicts the, Ψ represents the mean of $\sum_{j=1}^{N} \Gamma_j$ samples over N observations. Γ_j denotes individual data values, Φ_j indicates deviation from mean Ψ, CV_j is transformed output, V_j represents weighting factor, and V_j denotes feature vector contributing to final computation. The mathematical representation Equation (7–8).

$$CV_j = v_j V_j \tag{7}$$

$$X_{jl} = V_l^S \phi_j \tag{8}$$

CV_j denotes the computed coefficient value for the j^{th} feature, where $v_j$is its weight and $V_j$represents feature magnitude. X_{jl} is the transformed feature, V_l^S indicates selected subset values, and $\phi_j$denotes the feature mapping factor. However, the PCA optimizes low-dimensional polymer feature spaces, which improves prediction accuracy, eliminates redundancies, and strengthens structure-property relationship learning.

Proposed Electric Fish Optimizer-Driven Intelligent Support Vector Machine (EFO-ISVM)

After the process of feature extraction, the proposed predictive model combines the EFO algorithm and the ISVM method to establish an accurate nonlinear relationship between the structure and properties of polymers. EFO optimizes the hyperparameters in the best method possible through the process of exploration and exploitation based on the electric field detection mechanism. ISVM learns structural information and predicts physical and thermal properties of polymers with high generalization ability.

Intelligent Support Vector Machine (ISVM) for Polymer Structure–Property Relationship

SVM is a type of supervised learning technique for both classification and regression tasks through the generation of an optimum hyperplane for maximum separation of data samples. Therefore, traditional SVM suffers from some disadvantages such as the choice of kernel and parameters, high computation complexity when dealing with large datasets, and deprived results when it comes to highly nonlinear data environments. However, ISVM was developed to address these problems.

Intelligent Support Vector Machine (ISVM)

It is an intelligent form of SVM which employs adaptive and weighted learning methods to accurately model non-linear functions in the prediction of relationships between polymer structure and its properties. Nonetheless, ISVM suffers from high complexity, is sensitive to noise in input data and performs when dealing with large datasets of polymers. The application of ISVM in this research is aimed at achieving reliable mappings between polymer structural attributes and its properties.

$$\min_{x,a,\xi} \frac{1}{2}||x||^2 + D \sum_{j=1}^{m} X_j\left[1 - z_j e(w_j)\right] \tag{9}$$

$$\sum_{j=1}^{m} b_j z_j = 0 \tag{10}$$

$$T.s: 0 \le b_i \le CW_j, for\ j = 1, \ldots, m \tag{11}$$

Where Equation (9), $\min_{x,a,\xi}$ denote the objective function is minimal, $\frac{1}{2}||x||^2$ is weight vector, a is as, are slack variable, D is penalization value, X_j is training sample, z_j is target variable, and $e(w_j)$ is output function for predicting. Where Equation (10), b_j is coefficient of feature, z_j is feature, m is number of total variables, the summation should equal to *0*. Where Equation (11), $T.s$ denote constraints, *0* is backoff lower bound, b_i is value of backoff, CW_j is contention window value, m is

the number of total nodes. Although, the ISVM accurately predicts polymer properties from structural descriptors, improving nonlinear relationship modeling, generalization performance, prediction stability, and material design reliability.

Electric Fish Optimizer (EFO) for Adaptive Hyperparameter Tuning and Enhanced Prediction Performance

EFO efficiently tunes ISVM hyperparameters to improve nonlinear polymer property prediction accuracy, stability, convergence, and generalization capability. The EFO divides individuals into active and passive electric fish according to electric field frequency. In this research, EFO efficiently searches optimal solutions to enhance polymer structure–property relationship prediction accuracy and stability.

Population Initialization

The initial positions, frequencies, and amplitudes of the electric fishes in the optimization problem space are generated randomly.

$$e_j^s = e_{min} + \left(\frac{fit_{worst}^s - fit_j^s}{fit_{worst}^s - fit_{best}^s}\right)(e_{max} - e_{min}) \tag{12}$$

$$B_j^s = \alpha B_j^{s-1} + (1-\alpha)e_j^s \tag{13}$$

In Equation (12 and 13), e_j^sdenotes electric intensity of the j^{th}solution at iteration s, $e_{min}$and e_{max}represent minimum and maximum electric bounds, fit_j^s, fit_{best}^s, and fit_{worst}^sindicate current, best, and worst fitness values, B_j^sdenotes updated behavioral position, and αrepresents adaptive weighting coefficient.

Active Electric Field Localization

An active electric fish performs local search with the help of its electric field, Brownian motion, and neighborhood interaction. The mathematical Equation representation in (14 and 15).

$$q_i = (w_{maxi} - w_{mini})B_j \tag{14}$$

$$w_{ji}^{cand} = w_{ji} + \varphi q_j \tag{15}$$

q_i denotes search range magnitude, $w_{maxi}$and w_{mini}represent maximum and minimum weight limits, $B_j$indicates behavioral coefficient, $w_{ji}^{cand}$is candidate weight, w_{ji}denotes current weight, ϕrepresents adaptive scaling factor, and $q_j$indicates perturbation quantity.

Passive Electric Field Localization

A passive electric fish conducts global search in the population with the help of sensing the active electric fishes, position update, and diversity maintenance.

$$o_l = \frac{\frac{B_l}{c_{jl}}}{\sum_{i \in M_B} \frac{B_i}{c_{ji}}} \tag{16}$$

$$w_{ji}^{new} = w_{ji} + \varphi\left(w_{qi} - w_{ji}\right) \tag{17}$$

In Equation (16 and 17), $o_l$denotes selection probability, $B_l$and $B_i$represent fitness values, $c_{jl}$and c_{ji}denote distances between solutions, $M_B$indicates candidate set, w_{ji}^{new}represents updated position, ϕ is adaptation coefficient, and w_{qi}, w_{ji}denote neighboring and current solution vectors.

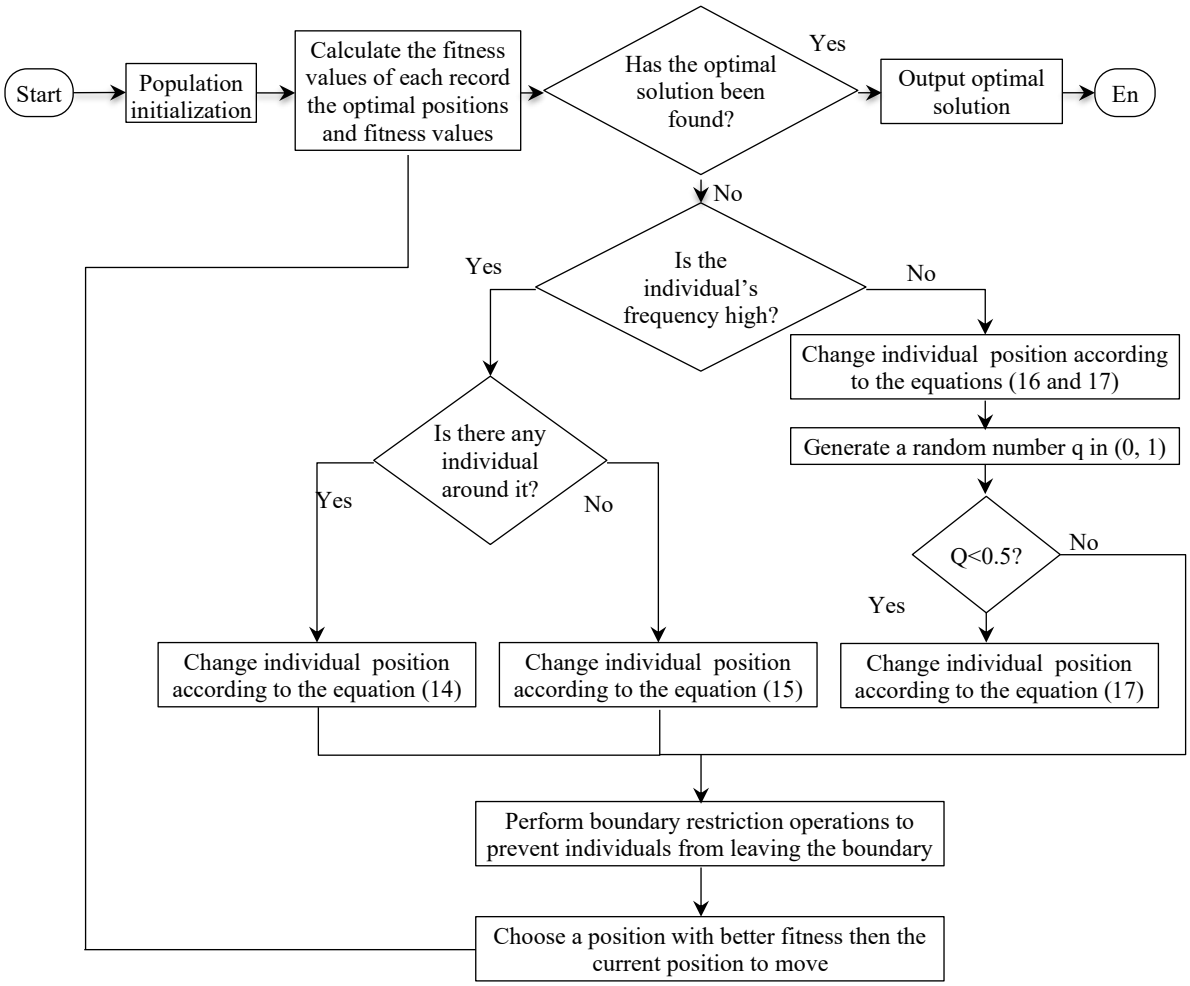


Figure 2. Process of electric fish optimizer.

The process flow of the EFO is shown in Figure 2, which includes initializing the population, localization of active electric fields, and passive localization. Active fish tend to migrate towards other active fish, while passive fish gravitate towards active fish groups, facilitating dynamic aggregation and optimization search capabilities. However, the EFO produces optimal ISVM hyperparameters, improving polymer property prediction accuracy, convergence stability, generalization capability, and error minimization. Algorithm 1 represents the entire process of EFO-ISVM.

Algorithm 1: Electric Fish Optimizer-Driven Intelligent Support Vector Machine (EFO-ISVM) for Polymer Structure–Property Prediction

Input:

Polymer dataset D

Population size M

Maximum iterations T

Structural descriptors S

Output:

Predicted polymer properties P ∗

Begin

1. Load and preprocess dataset D
2. Normalize data and handle missing values
3. Extract features using PCA
4. Initialize ISVM model

5. *Initialize Electric Fish population E*
For t = 1 to T do
  For each solution Ei in E do
    Generate ISVM hyperparameters
    Train ISVM using structural descriptors S
    Predict polymer properties:
    P = ISVM(S)
    Compute prediction error F(Ei)
  End For
For j = 1 to M do
    Perform active electric field localization
    Perform passive electric field localization
    Update candidate solution Ej
    Evaluate fitness F(Ej)
    If fitness improves then
      Update best solution Ebest
    End If
  End For
*Generate optimized prediction results P **
*Return P **
End

The EFO-ISVM model is a recently proposed hybrid algorithm that utilizes the EFO-based optimization for hyperparameters, ISVM based learning for adaptability, and PCA for structural feature extraction in the prediction of polymer structures and their properties. The EFO-ISVM can predict polymer properties with high precision, stability, and generalization while maintaining reliability in non-linear relationship mapping and structural integrity.

Structure–Property Quantification

The suggested model provides the ability to describe the relationships between the structures and properties of polymers through learning the relationship between the identified structural parameters and the desired properties of the polymers using the optimal predictive model. Such a process facilitates property prediction, selection of key structural factors, screening of materials, and intelligent polymer design.

EXPERIMENTAL SETUP

The experimental analysis of the suggested EFO-ISVM model is performed the high-performance computing environment to facilitate the processing of polymer datasets and robust learning. The implementation of the EFO-ISVM approach is done in Python using various libraries such as NumPy, Pandas, Scikit-Learn, SciPy, and Matplotlib. For the experiments, the following computing resources have been used: Intel Core i8 multi-core processor, 16 GB RAM, 1 TB SSD storage capacity, and Windows 11 operating system. Experimental findings are discussed along with an experimental analysis, comparative analysis with benchmark models, and validation of the model. Table 2 presents the evaluation metrics.

Dataset Partitioning

The data collected from polymers is then split into training and testing sets based on the ratio of 80% to 20%. The training set will be used in training and tuning the model, whereas the testing set will be used in assessing the prediction capability of the algorithm on novel polymers. To enhance robustness and mitigate overfitting issues, random shuffling and 5-fold cross-validation procedures are applied during the training phase.

The Polymer Structure–Property Dataset consisted of 5,000 samples. The dataset was partitioned into 4,000 samples (80%) for training and 1,000 samples (20%) for testing. To ensure reliable model evaluation and mitigate overfitting, 5-fold cross-validation was conducted on the training set, where each fold used approximately 80% of the training samples for model development and 20% for validation.

RESULTS AND DISCUSSION

The performance of the proposed hybrid approach combining EFO-ISVM was evaluated using quantitative performance measures and comparative analysis. The obtained results demonstrated improved prediction accuracy, enhanced generalization capability, reduced error tendencies, and faster convergence for polymer structure–property relationship quantification.

Experimental Analysis

Figure 3 illustrates the structure and thermal property relationships of the polymer via correlations analysis. The variability indicates the weak and independent relations between the descriptors, so the feature variance is retained in the predictive learning to quantify the structure-property efficiently and enhance the intelligent polymer prediction and characterization ability.

Figure 4 shows the decomposition results of the variation in the processing temperature to the trend, seasonality and residuals as time increases. The irregular patterns display the time-varying behavior in thermal processing characteristics, which implies the cyclic changes in polymer processing temperature. As one is aware of temporal behavior and enhanced the reliability of intelligent material processing.

Figure 5 (a) shows the variation of the thermal conductivity of the polymer to medium and high mechanical strength categories where shows irregular behavior in terms of thermal conductivity and moderate distribution for the features. Figure 5 (b) show low and high mechanical strength category of polymer where it is showing distinct change of conductivity and thus supports the relations between thermal and mechanical properties of polymer material.

Figure 6 displays density variations, and multidimensional structure and processing property relationships of the polymer material. The distributions reveal non-uniform material behavior and complex interaction patterns between structural, thermal and mechanical features. This may enhance the performance of predictive material investigation and support efficient design of intelligent polymers by exploiting multidimensional information of features.

Table 2. Performance evaluation metrics in polymer structure–property prediction model.

Metrics	Explanation
Mean Absolute Error (MAE)	Represents the average absolute difference between actual and predicted polymer property values, indicating prediction precision and error minimization capability.
Root Mean Square Error (RMSE)	Evaluates the magnitude of prediction errors by assigning higher penalties to larger deviations, reflecting model stability and regression effectiveness.
(R^2) Score	Measures the strength of correlation between predicted and actual polymer properties, indicating how effectively the model explains structural variance and property relationships.
Prediction Stability (%)	Indicates the consistency of prediction performance across multiple experimental runs and dataset variations, ensuring reliable generalization and reduced sensitivity to fluctuations.
Training Stability (%)	Measures consistency of model performance across different folds or runs, indicating low variance and reliable convergence behavior.
Optimization Capability (%)	Represents how effectively the model finds optimal hyperparameters, reflecting efficiency of search process and predictive performance improvement.

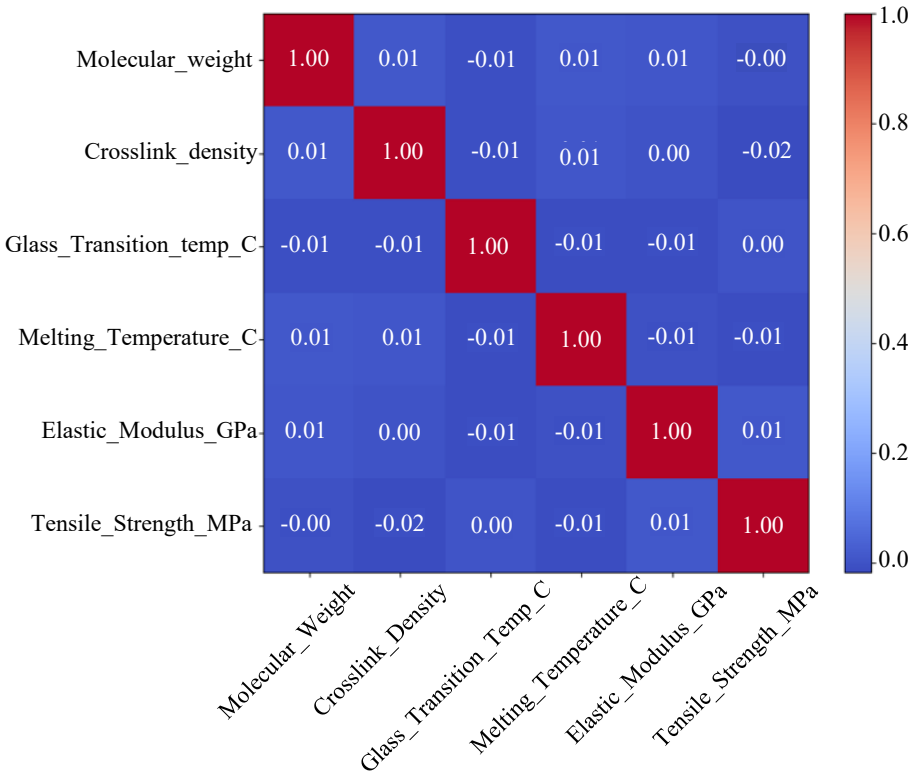


Figure 3. Correlation matrix of polymer material properties.

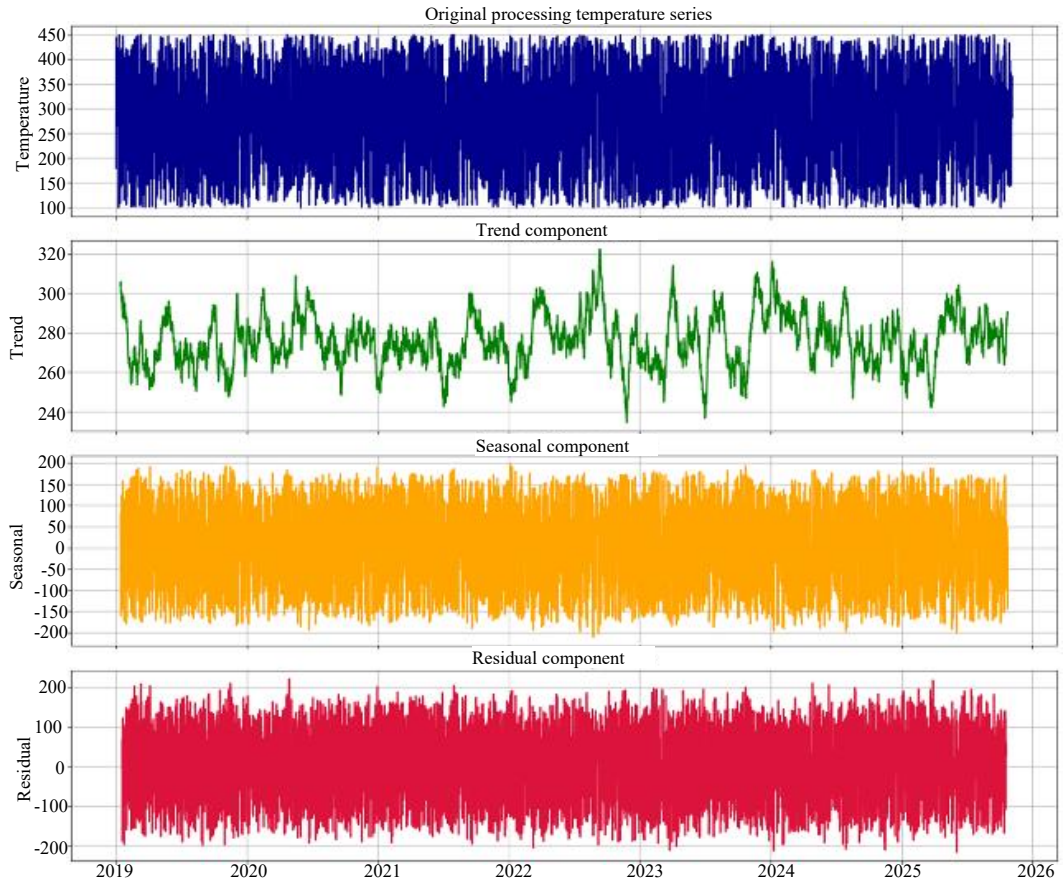
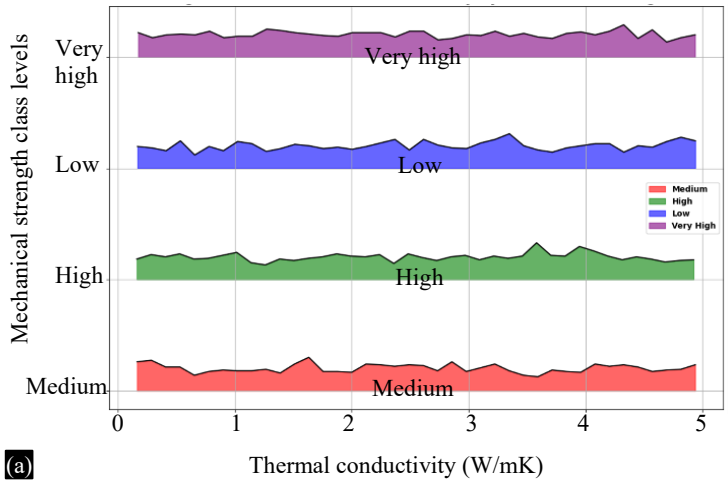
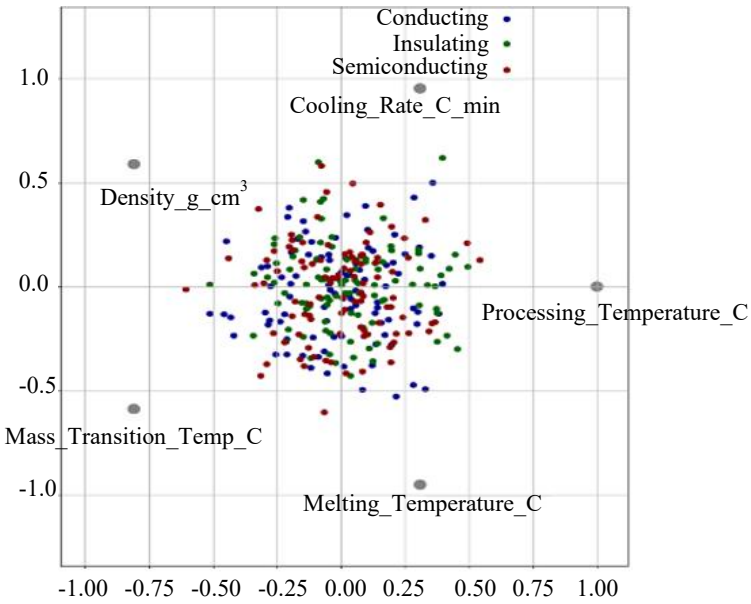


Figure 4. Time-series decomposition of processing temperature.



(a)



(b)

Figure 5. (a) Thermal conductivity by mechanical strength and (b) Visualization of polymer thermal characteristics.

Figure 7 (a) shows density distributions between the molecular weight and elastic modulus, where shows varying feature density for various polymers and clustered groups are non-uniform. Figure 7 (b) illustrates the 3-D relationship between cooling rate, water absorption and UV resistance with the structural complexity of the polymer, the result show interactions of multiple features and useful for polymer structure-property quantification with the intelligent prediction.

Fold Cross validation

The 5-fold cross-validation technique tests the reliability of the suggested EFO-ISVM method through varying training and testing of the method on different splits of the dataset to enhance its ability to generalize well. The fold that performed optimally in terms of having the lowest error rate and the highest optimization effectiveness was Fold 5. Table 3 shows the best results include MAE = 0.016, RMSE = 0.023, Training Stability = 97.6%, and Optimization Capability = 98.3%.

Comparative Analysis

Comparative analysis of the EFO-ISVM-based model is performed based on the comparison with the available ML techniques that include Artificial Neural Network (ANN) [16], Random Forest (RF) [17], AdaBoost Regressor [17], XGBoost Regressor [17], Gradient Boosting Regressor (GBR) [17],

Table 3. 5-fold cross-validation performance of the proposed EFO-ISVM model.

Fold	MAE	RMSE	Training stability (%)	Optimization capability (%)	Error standard deviation
1	0.020	0.028	96.4	97.1	0.0042
2	0.018	0.025	96.9	97.6	0.0038
3	0.017	0.024	97.3	98.0	0.0035
4	0.019	0.026	96.7	97.4	0.0040
5	0.016	0.023	97.6	98.3	0.0032
Average Performance (Mean ± SD)	0.018 ± 0.0013	0.0252 ± 0.0016	96.98 ± 0.42	97.68 ± 0.48	0.0037 ± 0.0004
95% Confidence Interval (CI)	[0.017, 0.019]	[0.024, 0.027]	[96.56, 97.40]	[97.20, 98.16]	[0.0033, 0.0041]

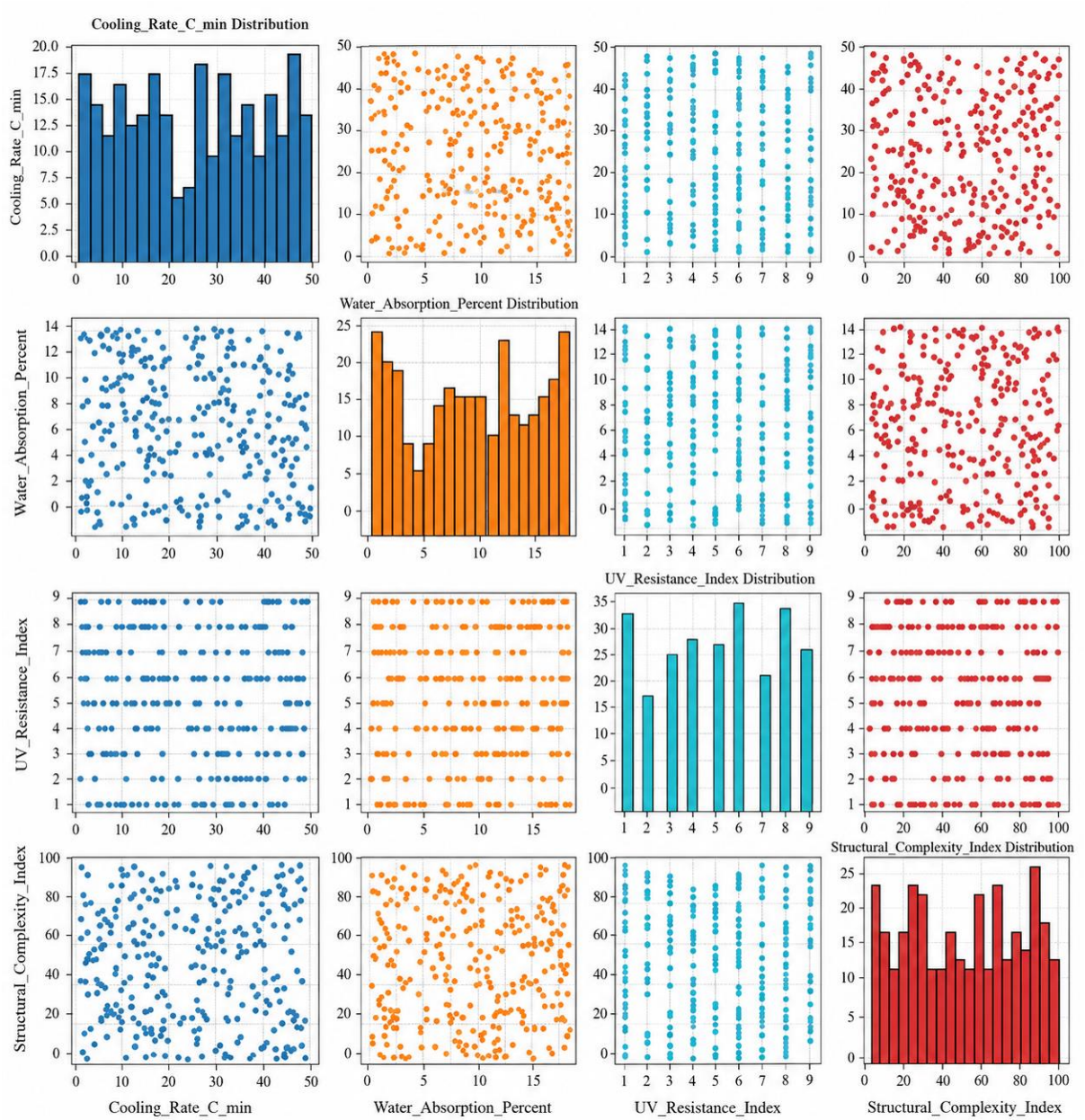
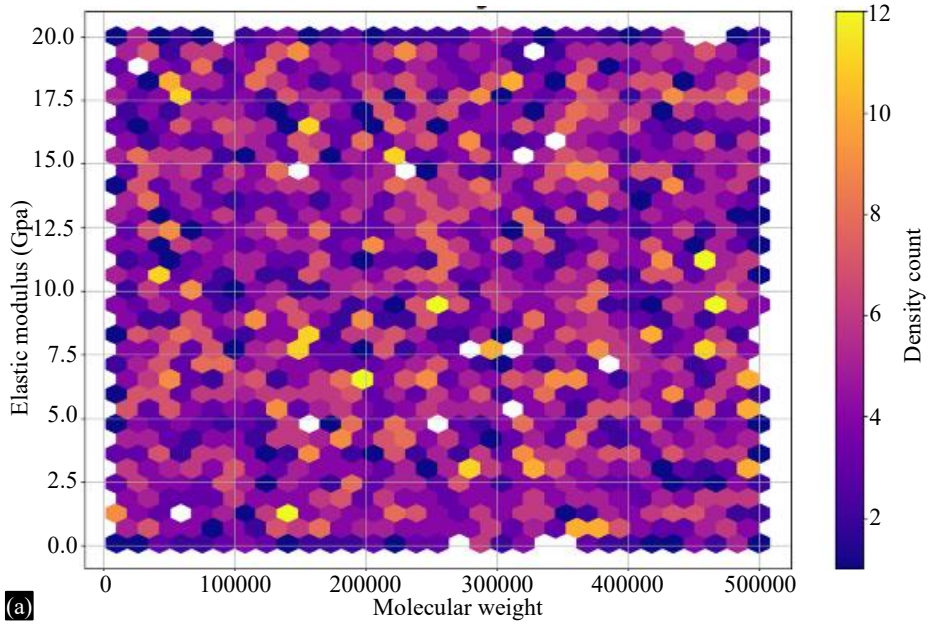


Figure 6. Polymer environmental and structure features.



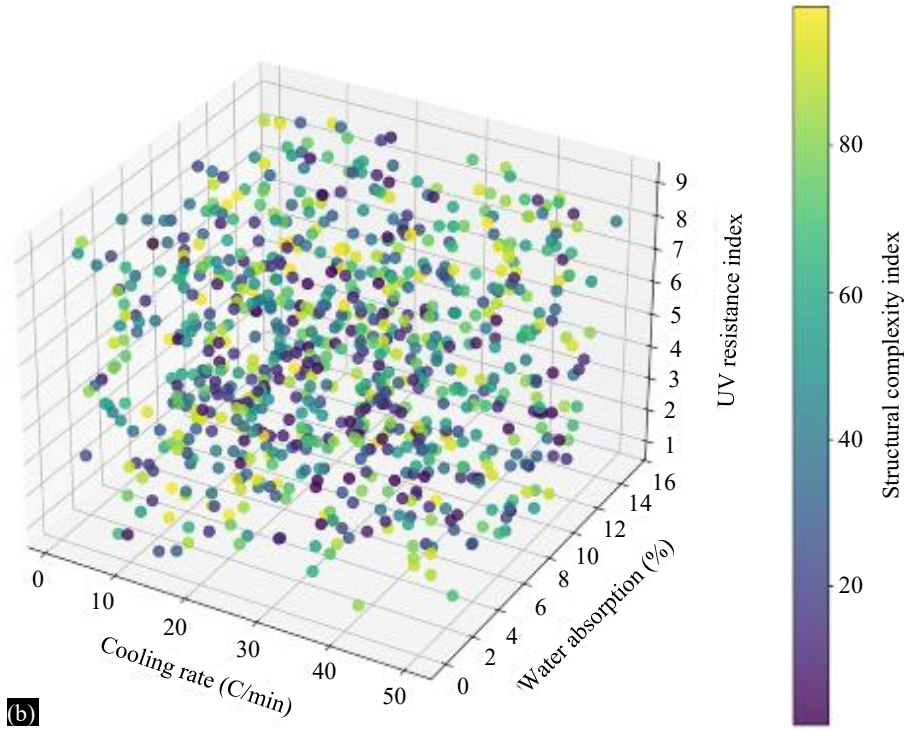


Figure 7. (a) Relationship between molecular weight and elastic modulus, (b) Key polymer structural indicators influencing material properties.

Linear Regression (LR) [17], Decision Tree Regressor (DTR) [17], and Bagging Regressor (BR) [17] for polymer structure–property estimation. All existing and proposed approaches have been trained and evaluated using the polymer structure–property dataset with same splits. Table 4 summarizes the comparative analysis of the above-mentioned approaches based on Accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2, and Prediction Stability.

Figure 8 show that the outcomes presented EFO-ISVM model has the best prediction performance in terms of having the highest prediction accuracy (97.84%), lowest MAR (0.018), lowest root MSE

(0.026), highest R^2score (0.964) and highest prediction stability (96.92%), implying better nonlinear structure-property learning performance. When compared to other models, XGBoost Regressor and GBR have fairly good performance in terms of accuracy of 96.08% and 95.17%, respectively, although they have comparatively higher error rates and low stability.

Structure–Property Quantification Insights

The developed model EFO-ISVM evaluates the correlation between polymer structural features and the fitness for material design. This evaluation is based on the contribution of critical thermal, mechanical, and structural factors. It has been shown that the percentage crystallinity, molecular weight, tensile strength, and crosslinking density have a substantial effect on predicting material behavior. High correlations mean enhanced thermal resistance, structure stability, and mechanical performance. Quantified correlations from the proposed model enable effective data-based polymer design and smart material selection. Table 5 Structure–Property Quantification Insights.

Table 4. Comparative performance analysis of polymer property prediction models.

Methods	MAE	RMSE	R^2 Score	Prediction stability (%)
ANN	0.046	0.061	0.923	92.11
RF	0.058	0.074	0.894	89.73
AdaBoost Regressor	0.052	0.068	0.907	90.64
XGBoost Regressor	0.027	0.038	0.963	95.21
GBR	0.031	0.043	0.951	94.38
LR	0.084	0.102	0.821	84.19
DTR	0.069	0.087	0.873	87.42
BR	0.039	0.054	0.936	93.07
EFO-ISVM [Proposed]	0.018	0.026	0.964	97.92

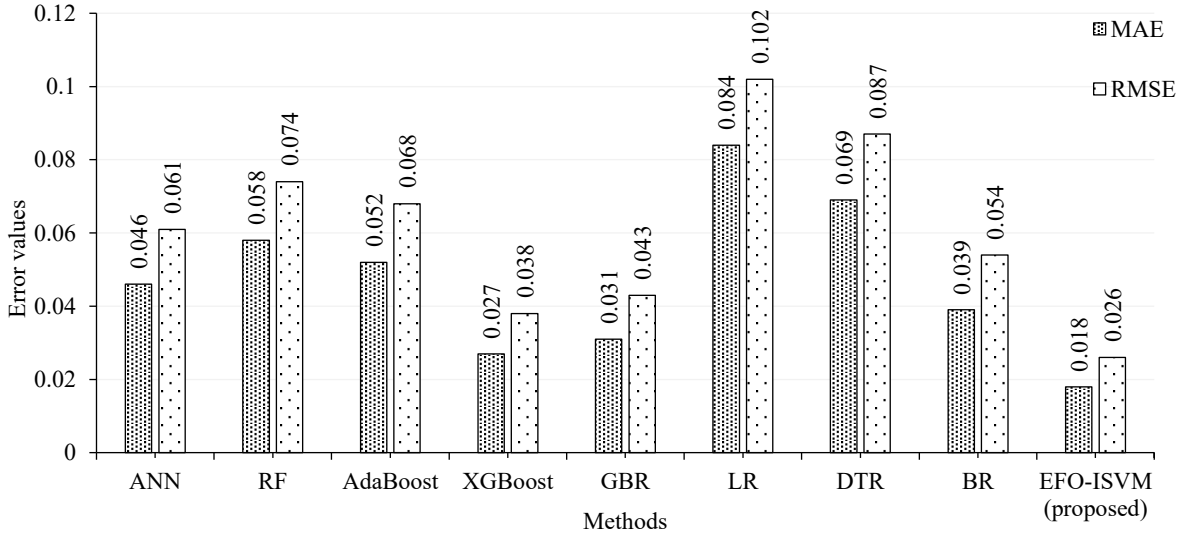


Figure 8. Comparison of MAE and RMSE across methods.

Discussion

The ML method was able to predict the polymer material structure via application of the EFO coupled with ISVM. Existing approaches that include the use of ANN [16] and RF [17] were considered as methodologies that are used to predict the polymer structure; however, it had no non-linear learning and generalization capabilities. Further, the conventional approaches encountered difficulties with respect to their inability to optimize efficiently and their failure to capture interactions in hierarchical form within the structural descriptors of the polymer structure.

Table 5. Structure–Property quantification insights.

Structural/Processing Feature	Related Property Behavior	Correlation Strength	Influence on Material Design Suitability
Crystallinity_Level	Mechanical rigidity and thermal resistance	0.91	SP
Molecular_Weight	Mechanical stability and chain integrity	0.88	SP
Crosslink_Density	Thermal stability and dimensional consistency	0.86	SP
Glass_Transition_Temp_C	Thermal operating capability	0.83	Positive
Thermal_Conductivity_W_mK	Heat transfer efficiency	0.79	Positive
Elastic_Modulus_GPa	Structural stiffness	0.87	SP
Tensile_Strength_MPa	Load-bearing capability	0.90	SP
Water_Absorption_Percent	Moisture sensitivity	-0.74	Negative
Cooling_Rate_C_min	Structural formation behavior	0.69	MP
Nano_Reinforcement	Composite enhancement	0.82	Positive
UV_Resistance_Index	Environmental durability	0.77	Positive
Structural_Complexity_Index	Molecular interaction diversity	0.73	MP

Note: *Strong Positive (SP), Moderate Positive (MP).*

This resulted in negative impacts on efficiency of predictions, generalizability, and material properties' estimation. In view of this, the suggested EFO-ISVM was regarded as an effective approach for the intelligent prediction of polymer structure and properties.

The proposed EFO-ISVM framework has strong potential for integration into industrial polymer design workflows. By enabling rapid prediction of material properties from structural and processing descriptors, the model can support material screening, formulation optimization, quality assessment, and accelerated product development. Its high prediction stability and adaptive learning capability make it suitable for deployment within digital manufacturing and polymer informatics platforms, reducing experimental costs and shortening material development cycles.

CONCLUSION

An ML model was applied to polymer structure and properties' relationship was introduced to provide predictive material design. The Polymer Structure Property Dataset, which is publicly available (Kaggle), was used in this research. The preprocessing of data involved the process of imputing missing values by using KNN and normalizing Z-scores. PCA was used for feature extraction. The novelty of this research is the use of the EFO-ISVM method for the tuning of parameters and nonlinear property estimation. The obtained results have shown higher prediction performance in relation to the mean absolute error MAE (0.018), root mean squared error RMSE (0.026), R^2 score (0.964), and prediction stability (97.92%) than the traditional ML methods. The proposed model include enhanced generalization capability, robustness in prediction, efficiency in non-linear transformation, and fast material property prediction. However, several limitations associated with the dependency of the model on the diversity of datasets and cost optimization. Future directions will involve integrating DL models, graph neural networks, molecular simulation properties, and autonomous materials discovery systems.

REFERENCES

1. Sharma, A., Mukhopadhyay, T., Rangappa, S.M., Siengchin, S. and Kushvaha, V., 2022. Advances in computational intelligence of polymer composite materials: machine learning assisted modeling, analysis and design. Archives of Computational Methods in Engineering, 29(5), pp.3341–3385. https://doi.org/10.1007/s11831-021-09700-9

2. Das, P.P., Rabby, M.M., Vadlamudi, V. and Raihan, R., 2022. Moisture content prediction in polymer composites using machine learning techniques. Polymers, 14(20), p.4403. https://doi.org/10.3390/polym14204403
3. Shireen, Z., Weeratunge, H., Menzel, A., Phillips, A.W., Larson, R.G., Smith-Miles, K. and Hajizadeh, E., 2022. A machine learning enabled hybrid optimization framework for efficient coarse-graining of a model polymer. npj Computational Materials, 8(1), p.224. https://doi.org/10.1038/s41524-022-00914-4
4. Conrad, F., Mälzer, M., Schwarzenberger, M., Wiemer, H. and Ihlenfeldt, S., 2022. Benchmarking AutoML for regression tasks on small tabular data in materials design. Scientific Reports, 12(1), p.19350. https://doi.org/10.1038/s41598-022-23327-1
5. McDonald, S.M., Augustine, E.K., Lanners, Q., Rudin, C., Catherine Brinson, L. and Becker, M.L., 2023. Applied machine learning as a driver for polymeric biomaterials design. Nature Communications, 14(1), p.4838. https://doi.org/10.1038/s41467-023-40459-8
6. Hooshmand, M.J., Sakib-Uz-Zaman, C. and Khondoker, M.A.H., 2023. Machine learning algorithms for predicting mechanical stiffness of lattice structure-based polymer foam. Materials, 16(22), p.7173. https://doi.org/10.3390/ma16227173
7. Shaari, N., Osman, S.H., Kamarudin, S.K., Karim, N.A., Basri, S. and Wani, A.A., 2026. Structure–property relationships in polymer electrolyte membranes for fuel cell applications: insights from molecular dynamics simulations. Polymer Bulletin, 83(5), p.243. https://doi.org/10.1007/s00289-026-06305-9
8. Huang, X., Ma, S., Zhao, C.Y., Wang, H. and Ju, S., 2023. Exploring high thermal conductivity polymers via interpretable machine learning with physical descriptors. npj Computational Materials, 9(1), p.191. https://doi.org/10.1038/s41524-023-01154-w
9. Kazemi-Khasragh, E., Blázquez, J.P.F., Gómez, D.G., González, C. and Haranczyk, M., 2024. Facilitating polymer property prediction with machine learning and group interaction modelling methods. International Journal of Solids and Structures, 286, p.112547. https://doi.org/10.1016/j.ijsolstr.2023.112547
10. Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V. and Borodulin, A., 2024. A multi-objective optimization of neural networks for predicting the physical properties of textile polymer composite materials. Polymers, 16(12), p.1752. https://doi.org/10.3390/polym16121752
11. Struble, D.C., Lamb, B.G. and Ma, B., 2024. A prospective on machine learning challenges, progress, and potential in polymer science. MRS communications, 14(5), pp.752–770. https://doi.org/10.1557/s43579-024-00587-8
12. Alhusban, M., Alhusban, M. and Alkhawaldeh, A.A., 2024. The efficiency of using machine learning techniques in fiber-reinforced-polymer applications in structural engineering. Sustainability, 16(1), p.11. https://doi.org/10.3390/su16010011
13. Qiu, H. and Sun, Z.Y., 2024. On-demand reverse design of polymers with PolyTAO. Npj Computational Materials, 10(1), p.273. https://doi.org/10.1038/s41524-024-01466-5
14. Berladir, K., Antosz, K., Ivanov, V. and Mitaľová, Z., 2025. Machine learning-driven prediction of composite materials properties based on experimental testing data. Polymers, 17(5), p.694. https://doi.org/10.3390/polym17050694
15. Malashin, I.P., Martysyuk, D., Nelyub, V., Borodulin, A., Gantimurov, A. and Tynchenko, V., 2025. Deep learning for property prediction of natural fiber polymer composites. Scientific Reports, 15(1), p.27837. https://doi.org/10.1038/s41598-025-10841-1
16. Ishikiriyama, K., 2022. Polymer informatics based on the quantitative structure-property relationship using a machine-learning framework for the physical properties of polymers in the ATHAS data bank. Thermochimica Acta, 708, p.179135. https://doi.org/10.1016/j.tca.2021.179135
17. Malashin, I.P., Tynchenko, V.S., Nelyub, V.A., Borodulin, A.S. and Gantimurov, A.P., 2023. Estimation and prediction of the polymers' physical characteristics using the machine learning models. Polymers, 16(1), p.115. https://doi.org/10.3390/polym16010115

18. Aditya Gujar, Manas Kale, Vaibhav Gobbur, Maya M. Charde, Amol J. Asalekar. Noise and Vibration Reduction Using Sustainable Polymer Nanocomposite Materials: A Review. Journal of Polymer & Composites. 2026; 14(03).
19. Vijay Kumar, Apurva Jain, Rachna Narula, Megha Sehgal, Meena Siwach, Tulika Bhatia. Modern Heritage Restoration Using GANs: Insights into Polymer–Composite Aging. Journal of Polymer & Composites. 2026; 14(02):508–516.
20. Gobind, Tejeet Singh. Fabrication and Characterization of Hybrid Composites Using Natural Fibers Reinforced in HDPE and Optimization of Injection Moulding Process Parameters to Minimize Shrinkage. Journal of Polymer & Composites. 2026; 14(02):556–585.
21. Priya R, M. Bala Theja, Hari kumar Andem, Ankush B. Khansole, S. Mohamed Rabeek, Rekha Anantharaman, S B G Tilak Babu. Synergistic Effects of Hybridized Nano-Silica and Hemp Fiber Reinforcement on Bio-Epoxy Composites. Journal of Polymer & Composites. 2026; 14(03).
22. Shashank S. Patokar, Bhagyashri P. Thakur, Sairaj S. Gujar, Maya M. Charde, Amol J. Asalekar. Natural Frequency Analysis of Polymer Composites. Journal of Polymer & Composites. 2026; 14(03).
23. G. Santhosh Kannan, M. HemaPriya, B. Vijaya, J. Faney. Mechanical Performance and Material Properties of Nsm CFRP Polymer System. Journal of Polymer & Composites. 2026; 14(01):311–319.

