Preprints.org

Review

Recent Advances in Polymer Design through Machine Learning: A Short Review

Bita Yarahmadi and Seyed Majid Majid Hashemianzadeh [*]

Review

Recent Advances in Polymer Design through Machine Learning: A Short Review

Bita Yarahmadi and **Seyed Majid Hashemianzadeh** *

Molecular Simulation Research Laboratory, Department of Chemistry, Iran University of Science and Technology, Tehran, Iran

* Correspondence: hashemianzadeh@iust.ac.ir; Fax: +982177491204; Tel: +9821772402871

Abstract

Polymers are widely used across diverse industries, including medicine, energy storage, construction, aerospace, agriculture, transportation, and electronics. However, the complexity and variability of polymer chemical compositions and structures present significant challenges for their development. The integration of machine learning (ML) algorithms with large datasets has opened new avenues for advancements in polymer science. Polymer informatics focuses on predicting polymer performance and optimizing synthesis conditions using ML models. With the growing availability of comprehensive datasets and ongoing improvements in ML techniques, polymer informatics can now be applied more effectively and systematically. This study provides a concise overview of the application of various ML approaches, including supervised learning, unsupervised learning, and artificial neural networks (ANNs), for predicting and classifying the properties of different polymers.

Keywords: polymers; machine learning; artificial neural network; regression algorithm; classification algorithm

1. Introduction

Polymers play a fundamental role in materials science with a wide range of applications, including the production of everyday products such as plastic packaging, lithium-ion batteries, 3D printing materials, and solar cells [1]. In polymer research, researchers encounter the challenge of high-dimensional data and an extensive body of literature, encompassing numerous studies on critical variables such as monomer diversity, varying polymer synthesis conditions, and intricate chain configurations. These factors, coupled with the inherent limitations associated with managing high-dimensional data and synthesizing information from a vast number of publications, present substantial challenges within this domain [2]. Polymer research is often constrained by insufficient

data, with the field predominantly relying on empirical "trial and error" approaches informed by extensive experimentation and accumulated experience. This methodology is inherently inefficient and slows down innovation in the development of advanced polymer materials [3].

Recently, with the help of various algorithms and extraordinary computing power, ML(ML) has evolved and become very efficient in processing high-dimensional data [4]. The fast adoption of ML with powerful regression and classification capabilities has opened new possibilities in many fields, including polymers [5]. The application of various ML techniques in the field of polymers is shown in Table 1. However, the development of polymer informatics is in its infancy. More research is underway to continuously accumulate data, optimize ML algorithms, and deeply integrate these factors in specific applications [6]. Each of the three main methods of ML including unsupervised, supervised, and reinforcement learning, has several diverse types of algorithms that are used depending on the research issue, helping to analyze operational data and reduce costs. Learning begins with observation and identifies patterns by experimenting with them [7]. Hundreds of new algorithms are created and tested daily, classified based on learning style (e.g., unsupervised, supervised, semi-supervised) or sometimes on the similarity in performance. The primary goal of algorithms is to generalize insights beyond trained samples to interpret data. ML algorithms build mathematical models based on sample or training data to make forecasts and decisions without programming [8]. Researchers strive to develop new learning methods and explore their feasibility and learning quality. Simultaneously, some researchers are applying ML techniques to new problems, significantly reducing operational costs and speeding up data analysis. The scope of ML research is vast [9]. The purpose of this review is to provide an overview of the algorithms used in the field of polymers. To keep things concise, we will not delve into the principles of modeling, but instead offer a summary of ML algorithms, as these principles have been extensively discussed in literature. For readers looking to delve deeper into these issues, we recommend references [10–12].

Table 1. ML methods used in the field of the polymer composite.

Polymer Composites	Characterization	Optimization methods	Inputs	Outputs	Reference
Glass fiber polymer composites	Erosion behavior	Response surface methodology (RSM) and ANN	Nozzle diameter, slurry pressure, impingement angle	The rate of the erosion	13
Hybrid carbon–glass epoxy composites	Mechanical behavior	GLODS, ABAQUS	Ply fiber orientation	Displacement, Stress	14
Red brick dust- glass–epoxy composites	Wear behavior	ANN	RBD content, velocity, impingement angle, erodent temperature, erodent size	The flow of erosion	15
Silicon behavior carbide particle/glass fiber–reinforced Polymer	Machining behavior	Evolutionary computing	Inter-electrode gap, voltage, electrolyte concentration, and duty factor	The rate of material removal	16

matrix composites					
Hybrid fiber reinforced polymer composites	Tensile and inter-facial properties	Deep Neural Network (DNN) oppositional based Firefly optimization (OFFO)	Hybridization of the layer	Tensile, and failure strain, tensile strength, modulus	17

2. The ML Steps

The goal of ML is to improve the efficiency of computers in performing tasks over time. The ML process consists of significant steps including data collection, preprocessing of data, feature selection methods, algorithm selection, model training and data pipeline, model validation, model stability, and prediction [18]. The different steps of ML modeling are illustrated in Figure 1 and summarized below:
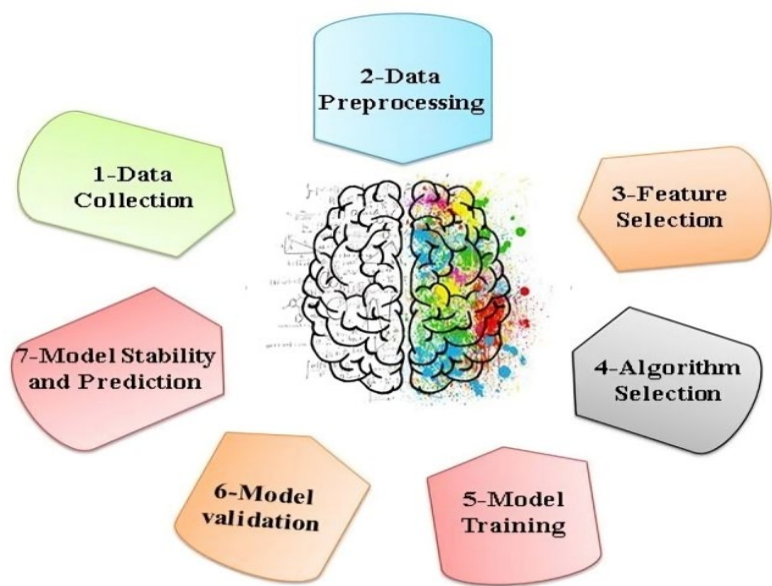
Figure 1. The important steps of ML modeling.

2.1. Data Collection

The procedure of extracting raw data for ML tasks is the first step in the learning process. The dataset is then divided into two groups, testing and training. Data sets play a vital role in predicting polymer properties. On the off chance that the information isn't collected carefully, the prepared model will be of low quality. Several polymer-related databases have been developed and improved, some of which are presented in Table 2. In addition, there is no database containing detailed information and synthesis of polymer analysis results. The polymer community should encourage

more scientists to share their data and code with the public. ML technology can use large and growing polymer databases to discover richer and more powerful rules in polymers [19].

Table 2. Examples of databases used in polymer science [20].

Name	Description	URL
Protein Data Bank (PDB)	3D structures of complex, nucleic acids, and proteins	http://www.wwpdb.org
PoLyInfo	different data needed for polymeric material design	https://polymer.nims.go.jp/
Physical Properties of Polymers	physical properties and characterization techniques of polymers	G. Wignall, J. Mark, L.Mandelkern , K. Ngai, W. Graessley, E. Samulski, and J. Koenig
Nano Mine	An open-source data resource used in nanocomposites	An open-source data resource for nanocomposites
Carination	Computed and practical properties of materials	https://citrination.com/
NMR Databases	NMR spectra of polymer	https://www.acdlabs.com/dbs/nmr_db
Material characteristic Database	Engineering material characteristic	https://www.makeitfrom.com/
NIST Synthetic Polymer MALDIR Database	Matrix-assisted laser desorption ionization (MALDI) mass spectrometry on a broad variety of synthetic polymers	https://maldi/http://nist.gov/
CROW Polymer characteristic Database	A multitude of polymer characteristic	http://polymerdatabase.com/

2.2. Data Preprocessing

preprocessing is a essential step after collecting relevant data to ensure data is in a correct format useful for training a ML model, especially when using algorithms that are sensitive to the magnitude of input values .One of the key aspects in data preparation is scaling data, typically achieved through two methods: normalization and standardization. Standardization involves transforming the data so that its meaning is zero (μ=0) and its standard deviation is one (σ=1). On the other hand, normalization transforms the data to fall within the interval [0, 1], ensuring a range of variation equal to one unit (R=1) [21].

2.3. Feature Selection Methods

These methods have become essential in data processing. Selecting the right features can enhance inductive learning. By removing variables that have a minimal impact on the output, a more efficient model can be obtained. Feature selection is defined as the process of identifying relevant variables and eliminating irrelevant ones iteratively to create a subset of features that effectively describe the problem. This step offers several advantages, including improving the efficiency of ML algorithms, enhancing understanding of the data, gaining insights into the process, simplifying visualization, reducing overall data, potentially cutting costs, conserving resources in future data collection cycles or operations, using simpler models, and increasing speed [22].

2.4. Algorithm Selection

One of the biggest challenges in selecting appropriate and optimal algorithms lies in the explicit requirements of the problem at hand. There are numerous types of ML algorithms that have been developed and adapted to address various types of problems. Factors such as analytical purpose,

accuracy, speed, and unbalanced disadvantages play a significant role in determining the most suitable algorithm. The purpose of the analysis and the nature of the target factor are the key considerations when choosing algorithms and techniques. Each type of analysis - descriptive, predictive, and prescriptive - will require specific algorithms. Generally, different problems will necessitate different algorithms. For tasks where high accuracy is the goal, algorithms like Random Forest (RF), support vector machines (SVM), and artificial neural networks (ANN) are often recommended over others. For tasks where speed in modeling is more important, algorithms like logistic regression, linear regression (LR), and decision trees (DT) tend to outperform others. A significant challenge in modeling is dealing with imbalanced classes in the target variable. While methods like oversampling or under sampling can address this issue, certain algorithms like CART, C4, and C5 are also adept at handling such situations [23,24].

2.5. Model Training and Data Pipeline

In ML, models are utilized to learn patterns, relationships, and information within data and utilize them to predict target features. Once an optimal algorithm is selected, a model is trained upon the chosen algorithm and a data pipeline. The primary advantage of using pipelines is to prevent the leakage of training data into testing data and vice versa [25].

2.6. Model Validation

Following training, the effectiveness of the model is validated on a test dataset. This data should have the same underlying distribution as the training dataset but must be distinct data that the model has not encountered before.

2.7. Model Stability and Prediction

Finally, after training and validating the model's applicability, the hyperparameters of the model are adjusted, and the complete model is utilized for predicting or classifying polymer properties [26].

3. Type of the ML

The field of ML is primarily divided into three types: unsupervised, supervised, and reinforcement learning that is shown in Figure 2. The concept of supervised learning involves providing the algorithm with data that serves as correct answers. During the training phase of a supervised learning algorithm, the algorithm explores for patterns in the data that correspond to desired outputs. Once trained, a supervised learning algorithm can then classify or predict new inputs based on the patterns identified during training. The objective of a supervised learning model is to forecast the label for new input data [27].

In unsupervised learning, there is no target variable. The superlative example of this algorithm is the automatic grouping of a population. In this method, the system automatically divides the data into similar and equivalent groups based on common characteristics. The third type of algorithms classified as unsupervised algorithms are called reinforcement learning. In this algorithm, a machine (or controlling program) is trained to make specific decisions based on the current position and allowed actions (e.g. move forward, move back). Initially, the decisions may be random, but with each action, the system provides feedback or a score. The machine then learns from this feedback to determine whether the decision was correct and adjusts its behavior accordingly in similar situations. Markov decision processes are an example of this group of algorithms due to their reliance on previous states and behaviors. Neural network algorithms also fall into this category. The term "reinforcement" in the naming of these algorithms deals with the feedback stage that strengthens and improves the program and algorithm [28]. Each of these approaches has its own set of disadvantages, advantages, and depending on the problem at hand, one method may prove superior to the others. In the following section, we will explore examples of how various ML methods have been applied in polymer informatics.
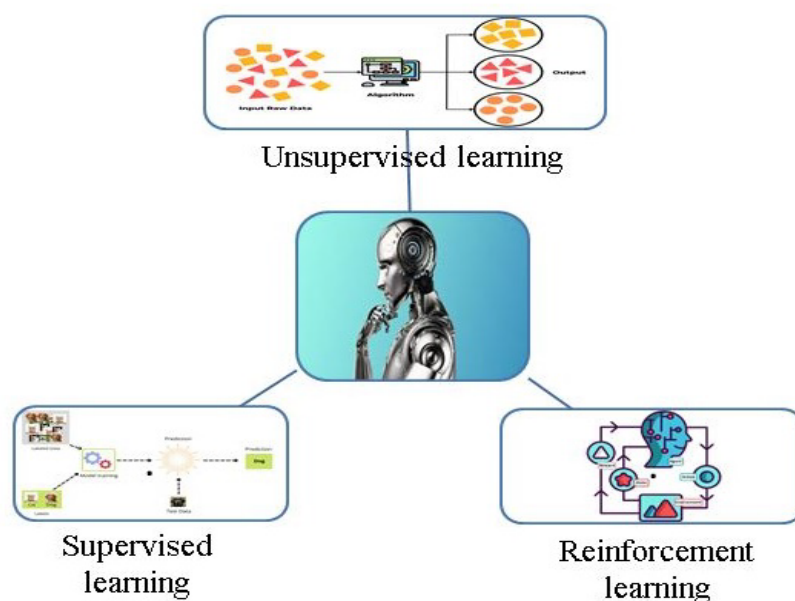
Figure 2. The difference between ML methods.

To determine which of the supervised or unsupervised algorithms is suitable for the problem, the following steps are used:

i.   Data evaluation: Are data labeled or not?

ii.   Specifying objectives: Is the topic repetitive and well-defined? Or is the algorithm needed to solve problems?

iii.   Cycle through available algorithm options: Is there a suitable algorithm to create a model with the highest accuracy according to the data distribution?

iv.   The big data classes can be a real challenge to monitor, but the results can also be very accurate and reliable. In contrast, large volumes of data can be managed in real time without supervision. However, there is a lack of clarity about the clustering method and the possibility of false results in this method [29].

3.1. Types of Supervised Learning Problems

The supervised learning method can be used to solve two categories of problems: 1. Regression, and 2. Classification problems. In each of these problems, we need labeled training data.

Classification is a method used in various ways to identify patterns and gain knowledge about the patterns governing different types of data. Regression involves supervised learning problems where the target values are continuous. In these types of problems, we have at least two variables: one being the independent variable (characteristic or input data of the model) and the other being the dependent variable (target value or output of the model). A comparison of the supervised algorithms that are used in polymer development is shown in Figure 3. In the following, the application of each of these algorithms in the field of polymers has been investigated [30].
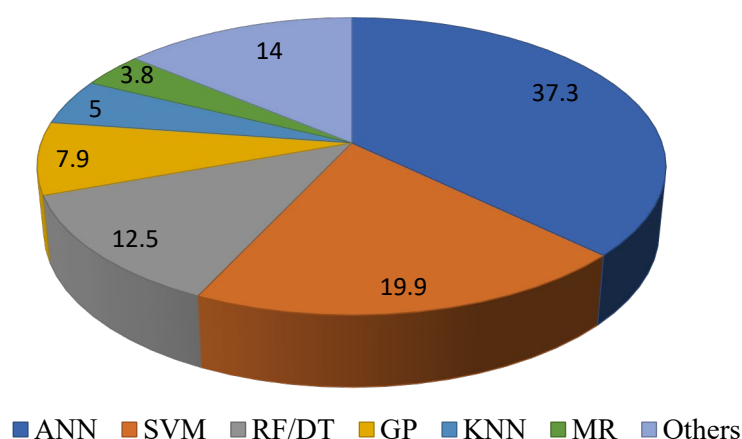
Figure 3. Comparison of the supervised algorithms that are used in polymer science.

3.1.1. Regression Algorithms

Regression is one of the important and used tools in the field of polymer property prediction [31]. The aim of a regression algorithm is to predict the output label or answer using continuous numeric values in the input data. Essentially, a regression model uses the properties of input data (independent variable) and continuous numerical output values (dependent variable or outcome variable) to learn the specific relationship between respective output and input [32]. Regression algorithms have garnered the attention of researchers for forecasting various properties of polymers. Lakshmi Y. Sujeeun and colleagues (2023) used six different supervised algorithms to create a model to classify and forecast the miscibility of polymer mixtures, including poly(hydroxybutyrate-covalerate)/fucoidan (PHBV/FUC), acetate/polyamide (CA/PA), and polyhydroxybutyrate/kappa-carrageenan (PHB/KCG). Their model inputs included physicochemical variables evaluated through Fourier transform infrared spectroscopy (FTIR), mechanical properties, and thermal analysis. Their results showed that the Random Forest (RF) algorithm executes better than other algorithms and that the regression model was related to the absorption of ultraviolet (UV) radiation by the polymer components [33]. Some regression algorithms like SVM and LR have fascinated the attention of scientists in the field of polymers [34]. If one independent factor is used to predict and identify the dependent variable, the model is called "simple linear regression". The form of the basic linear regression model is as follows:

$Y = \beta_0 + \beta_1 X + \epsilon$

This relationship is the equation of a straight line plus an error term, or ϵ. The variables of this linear model are the intercept (β_0) and the slope of the line (β_1). The slope of the line in simple linear regression indicates how sensitivity is for all factors. The intercept represents the value of the dependent factor and is calculated such that the value of the independent factor is zero. Alternatively, instead of removing the independent variable, a constant value or interception is considered as the mean value of the dependent variable. The LR algorithm is employed for predicting parameters like the miscibility of polymer mixtures, the impact of charge mixing, water mixing, additives, and temperature on polymer quality control, as well as the polymer framework based on monomers. The applications of the LR algorithm in the polymer field are shown in Figure 4 [35]. For example, Onur Balci et al. (2008) predicted wash fastness values of cleaned nylon 6.6 fabric dyed with 1:2 metal complex acid and CIE Lab data by using the ANN algorithm and LR models. They succeeded in predicting all the colorimetric data satisfactorily by using the LR model [36]. Sai-bing QIU et al. (2012) investigated the effect of charge mixing, water mixing, additive, and temperature on the quality

control of polymer-modified cement mortar using the LR analysis method. Parameters such as water-cement ratio and polymer-cement ratio as dependent variables, flexural strength, and compressive strength as independent variables of the regression equation showed multiple linear relationships between quality control of modified cement mortar and factors affecting the quality of cement mortar. Designing a polymer framework is a challenging problem that is easy to solve by using ML [37]. Also, Ryota Ido and colleagues (2021) used an LR algorithm to create a model to predict the polymer's framework as a function of the monomer and new descriptors to define the polymer's structure. Their results show that the proposed model can forecast the structure of the polymer framework containing up to 50 non-hydrogen atoms as monomers [38].
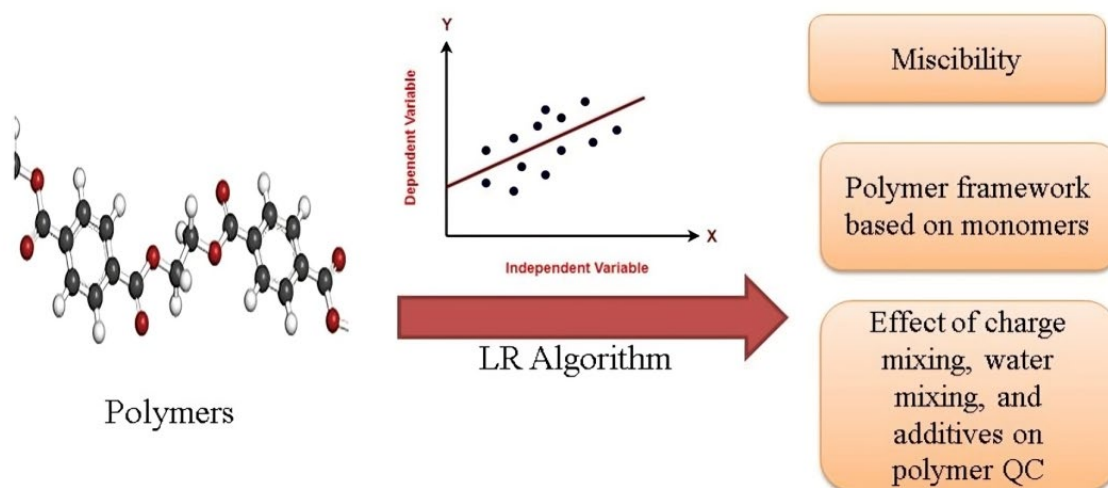

Figure 4. The applications of the LR algorithm in polymer development.

The goal of the SVR algorithm is to find a function that estimates the relationship between an input factor and a continuous target variable while minimizing forecast error. This algorithm has widely captivated the attention of researchers in the field of predicting the properties of polymers. The SVR algorithm has been utilized for predicting lubricant pressure for membrane separation, the abundance of polymer dielectrics at different frequencies, the molecular weight of biopolymers, determining the refractive index, and the energy gap of polymer composites. The applications of the SVR algorithm are shown in Figure 5 [39]. For instance, A.L. Ahmad and coworkers (2015) developed an approximate model to predict the plasticizing pressure for membrane separation of CO_2 using the basic theory of the two-state adsorption model. Fluidization pressure is a multivariate function of polymer properties. They used an SVR algorithm to correlate the fluidization pressure with the partial free volume and glass transition temperature of the polymer. To their results, the precise combination of these two characteristics allows the use of SVR to predict the plastic pressure of membranes used to remove CO_2 at high pressure from natural gas [40]. Yi, Y et al. (2021) generated a dielectric data set using a digital representation scheme. They used an SVM algorithm to create a model that predicts the abundance of polymer dielectrics at different frequencies (100 Hz to 1015 Hz) [41]. Senthil Kumar Arumugasamy et al. (2021) investigated the use of ANN and SVR algorithms to predict the molecular weight of biopolymers. They used the polymer polycaprolactone (PCL) synthesized using enzymatic catalysis. According to their results, using the SVM regression algorithm is a good method to predict the molecular weight of the PLC polymers [42]. Owolabi, Taoreed O, and colleagues (2021) used particle swarm optimization-based support ordinary linear regression (OLR) and vector regression (PSVR) models to determine the refractive index of polyvinyl alcohol composites. The n-PSV model estimated the energy gap and refractive index of composites using the energy gap as a descriptor and showed better performance than the n-OLR model [43]. Chen Choi et al. (2023) used a whale optimization algorithm (WOA) and multicore correlation vector

regression (MRVR) to successfully predict degradation rates in polymer electrolyte membrane fuel cells (PEMFCs). They used the WOA algorithm to automatically tune and optimize weights and kernel parameters to improve prediction accuracy and to automatically adjust and optimize the weight and kernel parameters to increase the accuracy of prediction [44].
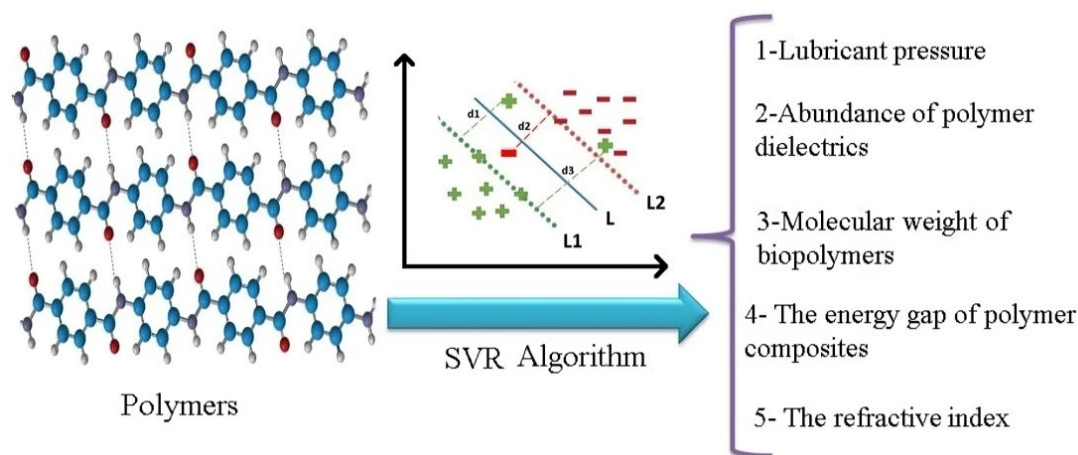

Figure 5. The applications of the SVR algorithm in polymer field.

3.1.2. Classification Algorithms

Classification algorithms belong to supervised learning, classify input data based on lineage, and are defined as the process of predicting classes or categories based on specified observations or points. Polymer classification results vary depending on the type of functional monomer, synthesis method, and conditions. Classification has the mathematical task of approximating a function (f) from input parameters (X) to output parameters (Y) [45]. Classification algorithms like SVM, k-nearest neighbors (KNN), logistic regression, and DT have fascinated researchers in the polymer field. For example, a good indicator of the efficiency of hollow carbon nanospheres is the structure of the core polymer particle [46]. Zhen Yang and colleagues (2013) used the SVM algorithm to optimize and classify the structure of core polymer particles. They succeeded in finding control criteria and optimum conditions for the pore structure of composite polymers [47]. Safwan Altarazi et al. (2019) used SVM algorithms to classify and predict the tensile strength of polymer films of various compositions as a function of processing conditions and considered two films manufacturing techniques including compression molding and extrusion blow molding. A combination of SVM and Genetic Algorithm (GA) enables highly accurate transformer fault diagnosis [48]. Yiyi Zhang and coworkers (2020) used GA-SVM and frequency domain spectroscopy (FDS) to predict insulation moisture content in oil-immersed transformers. They presented a method to construct a multiple GA-SVM classifier for moisture detection based on the fit analysis model [49]. In designing plastic products, it is important to adhere to product specifications based on needs and consider cost. Due to the wide variety of polymer materials available, selecting the appropriate raw material is a difficult task. Product properties determine whether the raw material is hard, flexible, elastic, etc. Javier Lorenzo Navarro et al. (2021) Classified plastics using hyperspectral images and various classical classifiers like SVM, 1D LSTM with RF, KNN, and CNN algorithms [50].

Researchers use unsupervised and supervised learning methods, contingent on whether the data is labeled or not. When some datasets have labels and others do not, a combination of two types of supervised and unsupervised algorithms is used. For instance, Xin Tian et al. (2022) presented a new method to find answer the problem of identifying weathered microplastics. They recorded the infrared spectra of diverse polymer particles (81,291 distinctive particles) using a quantum cascade laser (LDIR) and raw particle and weathered particle data using two supervised ML models including boosted decision trees (BDT) and subspace k-nearest neighbor (Sub-kNN), to analyze

labeled particles. They also used density-based spatial clustering for noisy applications (DBSCAN) models to group samples that could not be labeled using a supervised ML model. Their proposed method uses a combination of unsupervised and supervised learning models to significantly reduce the labeling effort [51]. Suresh Panchal and coworkers (2021) prepared polymer-modified quartz sensor arrays to analyze some volatile organic compounds like 1, 4-dimethoxy-2, 3-butanediol (BD), methanol, cyclohexanone (CH), acetone, and ethanol. They used three different algorithms, including linear discriminant analysis (LDA), DT, and RF, to classify the volatile organic compounds and their concentrations. Their model included three variables: reaction time, recovery time, and change in sensor resonance frequency [52]. Recently Hasan Dilbas and colleagues (2023) succeeded in designing a method to decrease the negative effects of superabsorbent polymers on the mechanical characteristic of concrete. Their method involved concrete mixing and processing and concrete mix design using DT algorithms, and they used air curing situations during the curing approach in preference to the standard water curing process [53].

Logistic regression is commonly used for classification problems with a discrete dependent variable. This algorithm distinguishes between two classes. Furthermore, this algorithm maximizes the probability of an event occurring and utilizes chi-square and Wald statistics to test the model's fit and the consequence of each variable's effect [54]. Logistic regression algorithm has been applied to classify polymers under different conditions [55]. Polymer composites are increasingly replacing metals in applications that require design flexibility, corrosion resistance, and a high strength-to-weight ratio. To reduce material costs and promote sustainable production, recycled materials are commonly used in the plastics industry. However, recycled materials have distinct properties, such as flow ability and viscosity, which can lead to quality issues. Julie Z. Zhang et al. (2014) developed a system to monitor burs on parts molded in a lab to simulate recycled plastics. They detected vibration signals during injection molding and used them as input factors in a logistic model to predict flash in the injection molding process. Detecting damage in polymer materials requires specialized techniques, as their damage types differ significantly from metals [56]. Ogheneovo Idolor and colleagues (2022) proposed an impair detection method for polymer composites that utilize naturally absorbed water. They employed a logistic regression algorithm to classify areas as "undamaged" or "damaged" to identify areas of damage. Selecting the appropriate framework material is crucial, along with other factors, when designing a suitable cartilage replacement material. Choosing the best polymer for cartilage repair is a costly and time-consuming process that often requires multiple trials [57]. Anusha Mairpady et.al (2022) introduced a reverse design attitude to predict the optimal polymer/blend for cartilage repair using a multinomial logistic regression (MNLR) algorithm. The MNLR algorithm identified polyethylene/polyethylene grafted poly (maleic anhydride) as the preferred material for cartilage patch [58]. The applications of the classification algorithms in polymer science are shown in Figure 6.
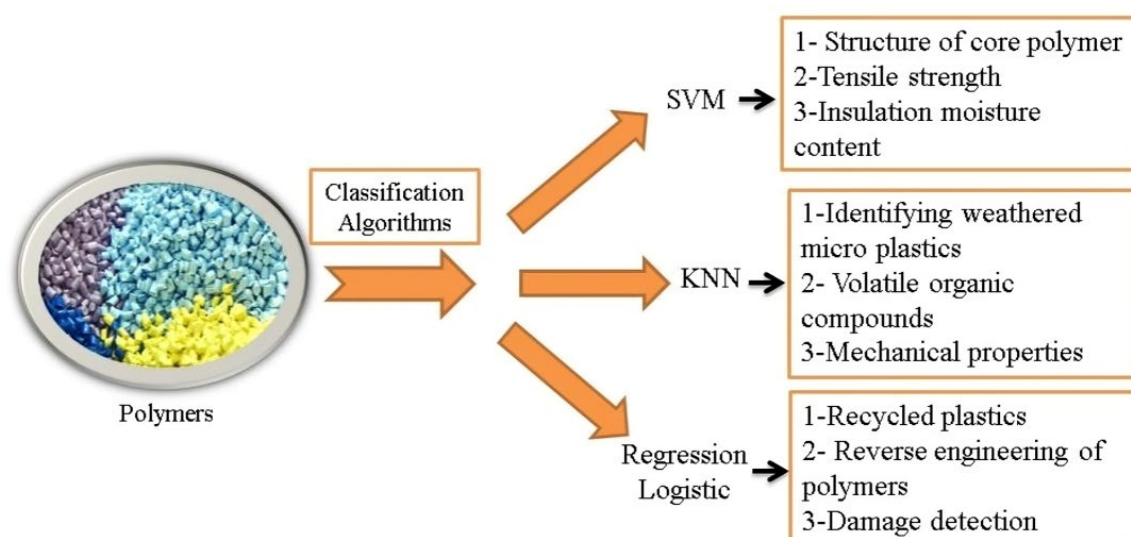
Figure 6. Some applications of classification algorithms in polymer science.

3.2. Unsupervised Learning

Unsupervised learning is an ML branch that does not require model supervision. Instead, models can work independently to discover previously undiscovered patterns and information, much of it in unlabeled data. In unsupervised learning, unlike supervised learning, the data is not specified in advance and connections between inputs and outputs are not the goal. This method has many advantages, including obtaining unlabeled data is easier and requires less intervention than labeled data, such as finding various unknown patterns in data or finding features that can help with classification [59]. Various studies have been published on the application of unsupervised methods in polymer informatics. For instance, Dazi Li et al. (2020) utilized cluster envelope modeling (CESM) to examine the clustering and shape behavior of PNC beads under stress based on MD simulation data. Simulations of matrix-free PNC-based elastomers and single-chain polymer nanoparticles that exhibited two types of deformation validated the accuracy of the proposed method. Yadava et al. (2015) introduced a method for selecting polymers for SAW vapor sensor arrays by applying a fuzzy C-means clustering (FCM) algorithm to the data. Their dataset included equilibrium partition coefficients for goal vapors in a wide range of potential polymers. The FCM algorithm groups polymers with similar characteristics into phase c clusters. Their results demonstrated that the proposed polymer selection method, prior to sensor fabrication, could enhance performance optimization and reduce the costs of electronic SAW sensors for various applications [60].

Another valuable algorithm for polymer classification is the k-means clustering algorithm, a quantitative method that divorce observations into k clusters based on the mean value [61].

In a study conducted by Lianbo Guo et al. (2018), laser-induced breakdown spectroscopy (LIBS) data was utilized to observe the behavior of engineered polymers in arm space. They employed ANN and K-means algorithms to classify engineered polymer species and enhance their performance. Initially, they classified 20 different polymers using the SOM neural network with adjusted spectral weights (ASW). The results indicated a successful separation of 18 different polymers, with polycarbonate (PC) and polystyrene (PS) being the only two polymers that were not separated. However, PC and PS were successfully classified using the K-Means clustering algorithm [62]. In a separate study by Danijela-Pesar and colleagues (2021), polymeric materials were classified based on properties such as brittle, hard, hard and strong, stiff and hard, and soft and soft using the K-means algorithm, yielding successful results [63]. Hiroki Kurita et al. (2022) utilized the k-means method to forecast Young's modulus and ultimate tensile strength (UTS) of carbon fiber reinforced polymer (CFRP), as well as practical values of Young's modulus and UTS. The effectiveness of the k-means method was demonstrated by accurately predicting UTS and Young's modulus of CFRPs with varying porosity and carbon fiber orientation [64]. Another category of unsupervised learning applicable in polymer development is dimensionality reduction algorithms. The primary purpose of reducing dimensionality is to prevent overfitting. Fewer variables in the training data lead to fewer assumptions made by the model, resulting in a simpler model. The advantages of reducing dimensionality include improved accuracy of the model by reducing misleading data, faster algorithms due to less computation, and less storage space required for the data. Dimensionality reduction is usually used along with other algorithms to reduce the modeling error [65].

Cross-linked polyethylene pipes (PEX-a) are a suitable alternative to conventional metal or concrete pipes. For the proper design and implementation of PEX-a pipe, it is important to determine the connection between the composition and the performance of the pipe. Melanie Hiles et al. (2019) used PCA, K-means, and SVM algorithms, along with data obtained from infrared spectral absorption peaks (IR) of PEX-a pipes, to create a model for classifying and comparing different formulas of PEX-a pipes. Using PCA results as input for K-means clustering and SVM led to significant error reduction and increased model accuracy [66]. Refractive index (RI) is a vital material property and is essential for proper material selection. Although accurate experimental measurements of RI are semi-empirical, time-consuming, and computational determinations of RI

are faster than empirical determinations. Jordan P. Lightstone and coworkers (2020) used measured RI data of polymers to create a model for near-instantaneous RI value prediction. They trained a Gaussian procedure regression model using data from 527 different polymers and used PCA to optimize model performance. The specific heat at constant pressure (Cp) of a polymer is a significant physical quantity that changes with temperature and is an essential parameter to describe the thermal conductivity of a material. Prediction of Cp presents major challenges [67]. Rahul Bhowmik and colleagues (2021) used decision tree methods, PCA, and took the first steps toward predicting Cp at room temperature. According to the results, their approach is useful for the development of new polymers with desirable Cp values. Due to the nanoscale dimensions, it is difficult to detect the surface area of polymer nanocomposites. Although electrostatic force microscopy (EFM) furnishes a way to measure local dielectric properties, extracting regional permittivity in complex surface geometries from EFM measurements is a challenge [68]. Praveen Gupta and colleagues (2021) performed interfacial permittivity extraction using data sets generated by simulating EFM artificial force gradient scans. To predict the permeability of polymers, they used two algorithms (SVR and RF), and PCA to increase the performance of the model. They succeeded in creating an accurate model to predict the permeability of polymer nanocomposites [69]. The unsupervised algorithms that have attracted the notice of researchers in classifying the polymers are shown in Figure 7.
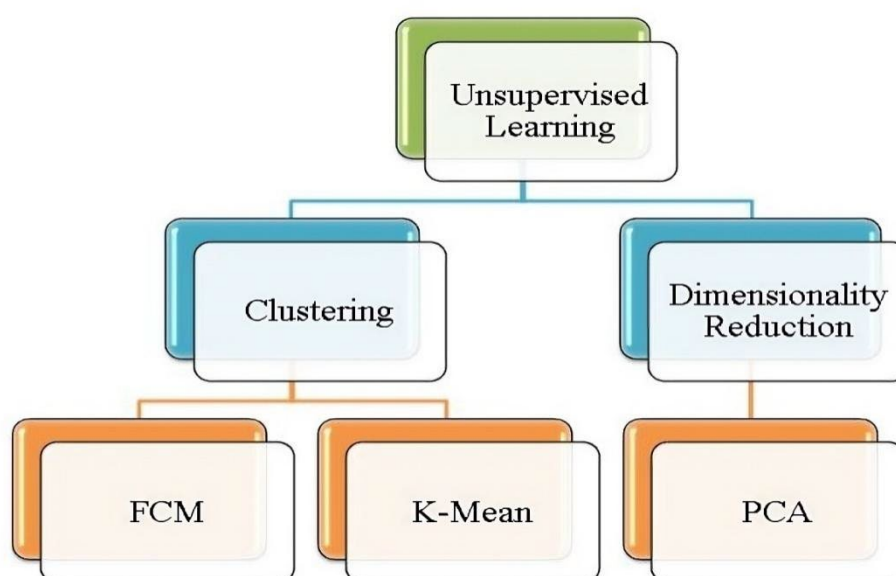

Figure 7. The unsupervised algorithms that have been used by researchers in classifying polymers.

3.4. Ensemble Algorithms

Ensemble methods combine the forecasts of multiple base estimators built using a specific learning algorithm to enhance universalizability and robustness compared to using a single estimator. These algorithms are powerful tools for various regression and classification problems, and their performance involves combining multiple weak outputs to achieve a strong final output. Using these algorithms reduces modeling errors and increases model accuracy. [70]. However, despite the improved accuracy, there are relatively few reports on the application of these algorithms in polymer design due to their computational complexity. For instance, Chenxi Liu et al. (2022) developed a sensitive fluorescent sensor based on dual molecular emitting polymers (dual-em-MIPs) for detecting perchloride chlorine in water and fish samples. They utilized an RF algorithm to predict the sensor response and analyze the fluorescence images, successfully predicting the sensor response with low error [71]. Among these algorithms, the Ada boost algorithm stands out for its high forecast accuracy and low error values when predicting or classifying polymer properties, garnering the

attention of many scientists. This algorithm used small decision trees. The tree should consider all training samples and give more weight to training data that are difficult to predict, while giving less weight to data that is easy to forecast. For a new input instance, each weak learner computes a predicted value as +1 or -1. The forecast for the group model is considered as the sum of the weighted forecasts. If the sum is positive, the first class is predicted and if it is negative, the second class is predicted [72]. Qichen Wang et al. (2022) utilized decision tree, Ada boost, and RF algorithms to predict the compressive strength of polymer composites. Their study revealed that ensemble algorithms outperformed single ML methods in predicting the compressive strength of the polymer composites [73]. Mahmoud Nasir Amin and colleagues (2022) employed the Ada boost algorithm for analyzing 481 samples with 9 input variables like curing time, temperature, sample age, NaOH molar concentration, Na_2SiO_3/NaOH ratio, and volatile alkalinity. They successfully developed a model to predict the ash content, aggregate volume, super plasticizers, and water in geopolymeric concrete mixtures [74]. Additionally, Madiha Anjum et al. (2022) created a precise model to forecast the compressive strength of fiber-reinforced nanosilica concrete (FRNSC) utilizing ensemble ML methods. They utilized four different algorithms, excluding gradient boosting (GB), RF, bag regression, and Ada boost regression. Their findings indicated that the Ada boost algorithm exhibited the highest accuracy and the least error in predicting CS values [75]. Devanshi Ledwani and coworkers (2022) developed models for predicting the melting flow rate of C2 and C3 polymers using LR, SVM, and Ada boost algorithms. Their results showed that the Ada boost regressor had the highest accuracy [76]. Recently, researchers have combined boosting algorithms with classification algorithms to enhance modeling accuracy. For instance, Wenjuan Sheng and colleagues (2023) utilized a distance least squares support vector machine (WLSSVM) and Ada boost algorithm with filter coefficients to differentiate various physical failure mechanisms in carbon fiber reinforced polymers (CFRPs). Filter coefficients were applied to eliminate unnecessary classifiers with minimal decision impact. Their results demonstrated that the improved Ada boost system with WLSSVM based on distance and filter coefficients outperformed advanced classification algorithms in distinguishing between different physical failures mechanisms in CFRP [77]. Also, the results of our research on predicting the amount of riboflavin absorption based on molecularly imprinted polymers showed that the use of the Ada boost algorithm has an excellent performance for predicting riboflavin absorbance [78]. Examples of polymer's properties that predicted using the Ada boost algorithms are shown in Figure 8.
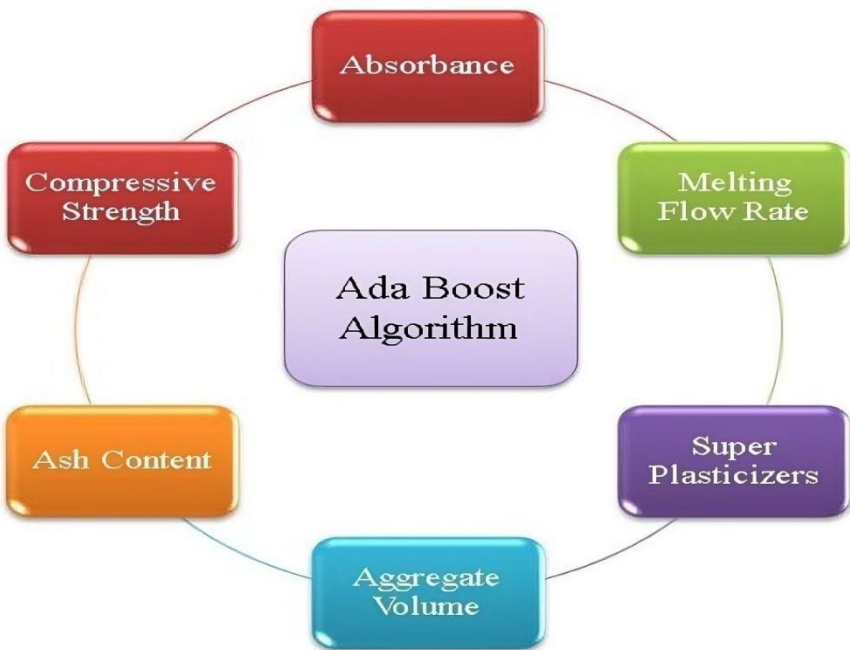
Figure 8. Examples of polymer's properties that predicted using the Ada boost algorithms.

4. Deep Learning

Deep learning was initially theorized in the 1980s; however, it did not gain significant attention until several years later. The reason for this attention is that deep learning requires large amounts of data to be successful. Greatest deep learning procedure utilizes neural network architecture, which is why deep learning models are often mentioned as deep neural networks. The term "deep" typically shows the number of hidden layers in a neural network. Traditional neural networks usually have only 2-3 hidden layers, whereas deep networks can have more layers [79].

Under the general title of "artificial neural networks" various types of computational models have been introduced, each used for a variety of applications inspired by specific aspects of the capabilities and characteristics of the human brain. These models involve a mathematical structure and are equipped with numerous parameters and adjustment mechanisms [80].

There are various reports on using neural networks to forecast different properties of polymers with high accuracy. Some properties of polymers predicted using the neural network algorithms are shown in Figure 9.
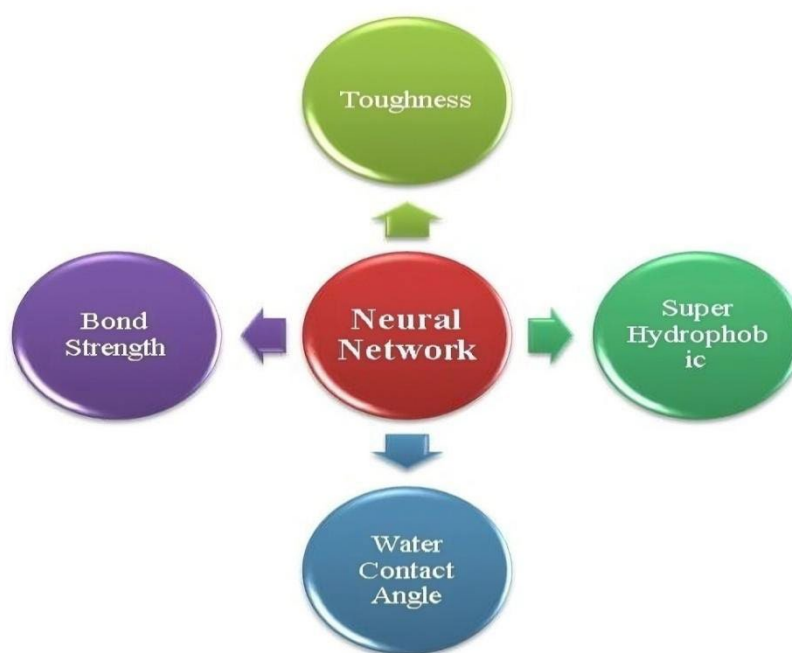

Figure 9. Some properties of polymers predicted using the neural network algorithms.

For example, Vinod Kushbaha and colleagues (2020) investigated the dynamic fracture toughness of silica-filled polymer composites under impact loading. They utilized a multilayer perceptron network to forecast the crack toughness of composite materials. The model's input parameters included velocity of the shear wave, loading rate, longitudinal wave velocity, time, volume fraction of silica filler, and a gradient descent function that was used to estimate the optimized synaptic weight. Good agreement was observed between the predicted and experimental values. According to their results, the strain rate over time is the most crucial factor when predicting the stress intensity factor. Predicting properties of non-synthetic polymers saves resources and time for industrial development and advances scientific understanding of structure-property affinity in polymer studies [81]. Qiang Wang et al. (2021) successfully developed a design map of polymer topography using experimental methods, numerical simulations, and ANN algorithms. The ANN model revealed various nonlinear relationships between topographic design parameters, and Laplace pressures, besides fundamental differences between micrometer and sub-micrometer length scales.

Their results showed that the ANN algorithm is a fast design tool for exploring surface topography. The repair effectiveness of fiber-reinforced polymer (FRP) depends on the bond strength between FRP and concrete [82]. Rami Haddad and coworkers (2021) used the ANN algorithm to predict the bond strength between FRP and concrete, utilizing data from the literature for their modeling. Their proposed model considers the influence of important parameters such as concrete compressive strength, adhesive tensile strength, maximum aggregate size, ratio of FRP to length and width of concrete, thickness of FRP, and modulus of elasticity on bond control. The model they proposed exhibited high predictive ability and adaptability on test data and training data, respectively, with predictive accuracy much higher than experimental models in the literature. Classification of fiber-reinforced polymer composites and textiles is a difficult task [83]. Nesrin Amor and coworkers (2021) successfully used ANN, genetic algorithms, and fuzzy logic to classify various problems in textile processing and fiber-reinforced polymer composites [84]. Convolutional neural networks (CNNs) are deep learning networks that learn directly from data. CNNs are especially useful for finding patterns in images to recognize classes and categories of polymers [85]. Minggang Zeng et al. (2018) used a GCNN algorithm to predict the dielectric constant and energy band gap of polymers. They showed that GCNN depends on the morphological data of polymers, eliminating the need for complex hand-crafted descriptors, while making fast and accurate predictions [86]. Liyong Ma et al. (2019) *evolved* an automatic blister defect detection system to assure the quality of Polymer Li-ion Battery (PLB) cell plates, instead of relying on manual detection. They used a CNN algorithm to detect blisters in cell plates based on images of the cell plates [87]. Also, Luis A. Miccio and colleagues (2020) were able to forecast the glass transition temperature of polymers using a CNN algorithm. They trained several networks with different parameters using a previously studied data set and applied the same method to an extended data set with larger Tg dispersion and polymer structure variations [88]. Recently, Göksu Taş et al. (2023) Successfully designed an experimental system to determine a dataset by measuring the flow parameters, voltage, and temperature in lithium-polymer batteries. They used this data set to evaluate the state of charge (SOC) of Li-polymer batteries and employed CNN as the method for predicting SOC [89]. In a study by Jia et al. (2025), deep learning models were used to design polymers with enhanced tear resistance. These models identified stress-sensitive molecules that could exhibit higher resistance to tearing, demonstrating the potential of deep learning to discover novel molecular features for improving polymer properties [90]. Moreover, in another study, Malashin et al. (2025) explored the application of reinforcement learning methods in polymer science. This study showed that reinforcement learning approaches can predict structure–property relationships, assist in polymer synthesis, forecast performance, and identify effective materials. These approaches, using machine learning models such as AdaBoost, Gradient Boosting, CatBoost, LightGBM, and XGBoost, improve prediction accuracy and manage complex data relationships [91]. The Information about the datasets, parameters, and architecture of some networks that are used in polymer science is shown in Table 3.

Table 3. Information about the datasets and architecture of networks that are applied in the polymers field.

Dataset	Input layers nodes	Hidden layer numbers	Output layer nodes	Activation function	Loss function1	Dataset size	Accuracy	Ref.
polymers and them their solvents	Two branches: Solvents and Polymers	Branch of solvent:2 Hidden Layers: 3	Binary classification: Good Solvent/NonSolvent	PReLU	NR.	4595 Test: 10% Training: 90%	%93.8	[92]
TC dataset for polymers	The list of integer identifiers of 250	1-3 Conv2., 1 pooling layer,	The structure of molecules	NR	RMSE5 and MAE6	Training: 90% Test: 10%	%94.79	[93]

		3FC. Layers						
polymers and their Tg value	Binary Images of SMILES information	2Conv layers, 1 Max Pooling, 1 FC	one neuron	ReLU	Median relative error	Train/test: 75/25 80/20 90/10	%94	[94]
dataset of the polymer genome project	two-dimensional graph	2Conv, 1 Pooling Layer	physical properties of polymers	The non-linear convolution functions	MAE	1073 Polymers Training:60% Validation:20% test:20%	MAE: 0.24	[95]
pressure probes footprint pictures	100*9 pictures	2Conv/2Max Pooling/3 FC	Five nodes	ReLU	MSE	3000 Images Train/test: 80%/20%	%99	[96]

5. Examples of ML Applications in Polymer Development

In general, the use of ML applications has various applications in polymer mixing including:

1- Modeling and Simulation: ML can be used to develop complex models of polymer mixing based on various factors. This allows for prediction of mixing behavior and optimization of the procedure [97].

2- Process Control: Controllers equipped with ML can effectively control and adjust mixing factors within predefined limits.

3- Error Detection: ML analyzes sensor data to identify and diagnose defects or anomalies in equipment during the mixing process. This allows for timely repairs and prevention of failure.

4- Recommender Systems: ML provides recommendations for optimal mixing based on polymer blend properties and offers technical specifications for the desired product. The applications of this method in polymer mixing processes are summarized in Table 4. Tires and composites are two polymer compounds that are very practical and have attracted the attention of scientists to find a way to facilitate their synthesis. In the following, the use of ML in their synthesis has been investigated [98].

Table 4. ML applications in modeling process factors [99].

Inputs	Algorithms	Outputs
Dwelling time, furnace temperature, tension applied to strands	ANN	final properties of carbon fibers
Reactor temperature /pressure, liquid level and flow rate of catalyst	DBN	Monomer converting, average of molecular weight, Reaction time, dispersion and heat stability
temperature, feed rate, reaction time and catalyst amount	SVM	Viscosity

Injection speed / pressure, duration of pressure, mold temperature, time Coolant, injection amount, screw rotation speed, body pressure, body temperature	ANN، GP, SVR	Process factors, fiber direction distribution, physical properties/ Mechanical, surface roughness
Feed rate, hydrogen concentration, and reaction temperature	PCA + GP	Prepare conditions and item quality
Material and operation factors (rotation speed, flow speed). output, temperature and composition	ANN، sPGDSVM، C2V, iDMD	properties and execution (Young's modulus, abdicate stretch, push)
process factors (position, contraction angle, polymer, channel width, and solvent flow	GP	Product characteristics (average length, average diameter, etc.)
Reaction conditions (temperature and initiator concentration)	ANN	Average molecular weight, monomer conversion, and viscosity Reaction mass

Tires

Rubber is a type of polymer with special physical properties such as flexibility, stretch ability, elasticity and durability. These selected properties of them have driven to a wide run of applications. Rubber products require specific properties such as wear resistance, toughness, hardness, and stiffness, which necessitate compounding rubber with varying amounts of additives [100]. On the other hand, the experimental method allows the designer to choose the initial combination that possesses the desired properties. While this combination may not always be the most cost-effective or optimal in terms of properties, the experimental method does benefit from the designer's experience [101]. Finding an accurate, fast, and cost-effective method for designing composite rubber compounds is crucial for the rubber industry. Rubber compound designers must consider the effects of each material on all properties and select the best combination based on price and specifications. It is important to pay attention to how raw materials interact with each other and how the design method impacts the final product [102].

The need of plan rules prevents a comprehensive understanding of the reverse relationship between mechanical behaviors and the plan handle. Traditional design strategies, such as structure-based design and experimental design, are time-consuming and require extensive database searches, especially for complex designs with multiple variables.

These strategies are constrained by computational costs and rely heavily on the designer's experience. Recent Advancements in ML methods have revolutionized material science, structural mechanics, and the optimization of composites and rubbers. ML algorithms can analyze big datasets to uncover configuration, develop mathematical relationships, and design new materials. ML technologies have the potential to innovate and optimize new applications by processing big amounts of data, learning patterns, and making algorithmic decisions to improve performance, efficiency, and overall quality [103].

Research has shown that using network modeling and synthetic compounds can predict the nonlinear behavior of rubber [104]. Correa and colleagues (2017) optimized three rubber formulations with an EPDM base to minimize costs. The results indicated that using a combination of MLmethods can quickly achieve optimal conditions at the lowest price [105]. Ghafarian and Hamedi (2020) conducted research to reduce time and costs, and increase accuracy in the rubber mixture design

process. The study focused on optimizing the rubber compound process by modeling the relationship between raw materials and properties using artificial neural networks. Each characteristic required a specialized neural network due to the irregular behavior and unique constants associated with each raw material. The results demonstrated that the rubber compound design process using this method was more effective than traditional experimental approaches [106]. In recent years several studies have explored the application of machine learning in tires, focusing on predicting tire lifespan, simulating wear under real-world conditions, and improving tire health monitoring systems. For instance, Karkaria et al. (2024) proposed a machine learning–based framework for predicting tire life and regenerability, concentrating on usage data instead of wear measurements [107]. Similarly, Jia et al. (2025) generated datasets of tire wear by simulating the wear process under real driving conditions [108]. Additionally, Tong et al. (2025) developed a machine learning–driven tire wear detection module capable of providing accurate results under laboratory conditions. These studies demonstrate significant advancements in using machine learning to enhance tire performance and safety [109].

D Printing of Composites

3D printing of composites has been the focus of decades of research, transforming them into essential components used across various industries and in design and production processes. These composites offer exceptional features such as a high strength-to-stiffness ratio, as well as significant weight savings. Additionally, they exhibit individual performance characteristics like dynamic response to external stimuli, making them ideal for intelligent systems made from materials like ceramics, common metals, or polymers, particularly with the introduction of Additive Manufacturing (AM) [110].

Through controlled 3D printing, structures can be created with complex geometric shapes and customized material attributes like stiffness, anisotropy, strength, and heterogeneity. The application of ML in 3D printing has led to the development of algorithms that enhance process optimization, manipulation, and customization. ML algorithms are utilized in AM to address factors such as local defects, internal stress, design accuracy, and microstructure changes. However, controlling these factors can be challenging due to the vast amount of data and variables involved. ML algorithms play a crucial role in understanding and improving the additive manufacturing process by developing patterns and models to reduce errors, minimize defects, and adjust microstructures. This involves rapid analysis, data collection, processing, and response control [111]. While statistical analysis can aid decision-making in certain cases, supervision of data history is necessary for optimization and customization processes [112].

6. Advantages and Disadvantages

There are many benefits to using ML, some of which are mentioned below:

1. ML helps identify trends and patterns, sift through big data and discover patterns that are invisible to humans.

2. Projects can be checked at each stage without the need for human intervention (automation). 3. Continuous improvement in system accuracy and efficiency occurs with time and experience and helps researchers make better decisions.

In fact, as the amount of data increases, the algorithm learns to make predictions more accurately and quickly.

4. The machine's algorithms perform very well in processing multidimensional and diverse data, even in dynamic environments. For all its benefits, power, and popularity, ML is not perfect [113].

Various laboratory analytical methods and theoretical calculations always have limitations, and ML methods are no exception to this rule. These limitations of ML methods are an undeniable subject that exists inherently in these methods, and researchers strive to reduce these factors. Some of these limitations are written below: 1. This method requires enough time for the algorithm to learn and develop its target accurately. 2. ML operations require huge resources. 3. Interpreting the results

poses a significant challenge, and the ability to accurately interpret the results produced by the algorithm is critical. 4. This method can have large errors in results due to human error. For example, when training an algorithm on a data set that is too small to be exhaustive (i.e., some populations are excluded), the result is biased predictions from the training set. Additionally, there are advantages and disadvantages to using any type of algorithm, which are presented in Table 5. The choice of the type of algorithm used for modeling has a major impact on the accuracy of the final model [114].

Table 5. Comparison of various ML algorithms [115,116].

ML algorithm	Advantages	Disadvantages
Linear regression	Simple method	Ignore of non-linear linkage between properties and descriptors
SVM, KRR	Cost-effective	The number of attributes and the size of the kernel matrix quadratically, resulting in large errors for huge datasets.
GPR	It predicts uncertainty for objective values	It requires a manageable data size and cannot train multiple features in a model
RF	This criterion has a specific measure for evaluating each descriptor and predicts large datasets with high accuracy.	It is possible that an overfitting problem will arise due to the generation of overly complex trees
ANN	It has a great ability to capture complex nonlinear relationships from huge data sets.	It cannot be interpreted, requires more data to train, and is time-consuming.
Deep neural network	Suitable for graphic representation of materials, learning different levels of abstraction levels	It cannot be interpreted, is time-consuming, and requires more data to train.

7. Conclusions

The use of ML has led to the development of polymers and significant progress in their synthesis. Different ML algorithms have been utilized to solve various problems, such as optimizing synthesis conditions, predicting the chemical and physical properties of polymers, and designing new polymers. The type of algorithm used relies on factors like the type of problem, including qualitative problems (such as classifying polymers based on specific properties like heat transfer coefficient) and quantitative problems (predicting factors like the refractive index of the polymer), as well as data type and data distribution. Various ML methods have been employed in polymer development, including unsupervised methods like clustering algorithms and supervised methods such as regression, classification, and ANN. Supervised algorithms commonly used in polymer research include regression logistics, SVM, LR, DT, ANN, CNN, and KNN. Unsupervised algorithms include PCA, k-means, and c-means. Regression algorithms predict numerical polymer property values, while classification and clustering algorithms categorize polymers based on their physical and chemical properties to solve qualitative problems. Ensemble algorithms, like Ada Boost, are highly accurate and minimize model errors. The accuracy of polymer property prediction or classification depends on factors like data quantity, data dispersion, algorithm type, use of feature selection methods, model accuracy enhancement techniques like pipelines, and hyperparameter tuning. Models created using various ML algorithms typically have high accuracy with minimal errors. ML modeling offers advantages such as cost reduction, time savings, increased accuracy, environmental protection, and elimination of chemical compound use. Accurate data collection is crucial, as errors can occur if data is not carefully collected. The use of ML in polymer research is an exciting field, and with advanced algorithms, accurate models will be created to predict properties, classify polymers, and discover new polymers in the future.

References

1. Jagadeesh, Praveenkumara, et al. "A comprehensive review on polymer composites in railway applications." *Polymer Composites, 2022,* 43.3: 1238-1251, https://doi.org/10.1002/pc.26478.

2. Ritsema van Eck, Guido C., Leonardo Chiappisi, and Sissi De Beer. "Fundamentals and applications of polymer brushes in air." *ACS Applied Polymer Materials, 2022,* 4.5: 3062-3087, https://doi.org/10.1021/acsapm.1c01615.

3. Sha, Wuxin, et al. "ML in polymer informatics." *InfoMat, 2021,* 3.4: 353-361,https://doi.org/10.1002/inf2.12167.

4. Ma, Ruimin, and Tengfei Luo. "PI1M: a benchmark database for polymer informatics." *Journal of Chemical Information and Modeling, 2020,* 60.10: 4684-4690, https://doi.org/10.1021/acs.jcim.0c00726.

5. Hatakeyama-Sato, Kan. "Recent advances and challenges in experiment-oriented polymer informatics." *Polymer Journal, 2023,* 55.2: 117-131, https://doi.org/10.1038/s41428-022-00734-9.

6. Chandrasekaran, Anand, Chiho Kim, and Rampi Ramprasad. "Polymer genome: a polymer informatics platform to accelerate polymer discovery." *Machine Learning Meets Quantum Physics*: 2020, 397-412, https://doi.org/10.1007/978-3-030-40245-7_18.

7. Schustik, Santiago A., et al. "Polymer informatics: Expert-in-the-loop in QSPR modeling of refractive index." *Computational Materials Science. 2021,* 194: 110460. https://doi.org/10.1016/j.commatsci.2021.110460.

8. Sahu, Harikrishna, et al. "An informatics approach for designing conducting polymers." *ACS Applied Materials & Interfaces, 2021,* 13.45 : 53314-53322, https://doi.org/10.1021/acsami.1c04017.

9. Patel, Roshan A., Carlos H. Borca, and Michael A. Webb. "Featurization strategies for polymer sequence or composition design by machine learning." *Molecular Systems Design & Engineering, 2022,* 7.6: 661-676, https://doi.org/10.1039/D1ME00160D.

10. Cai, Jie, et al. "Feature selection in machine learning: A new perspective." *Neurocomputing, 2018*, 300 , 70-79, https://doi.org/10.1016/j.neucom.2017.11.077.

11. Bertolini, Massimo, et al. "Machine Learning for industrial applications: A comprehensive literature review." *Expert Systems with Applications, 2021,* 175 , 114820, https://doi.org/10.1016/j.eswa.2021.114820.

12. Bikmukhametov, Timur, and Johannes Jäschke. "First principles and machine learning virtual flow metering: a literature review." *Journal of Petroleum Science and Engineering, 2020,* 184 , 106487, https://doi.org/10.1016/j.petrol.2019.106487.

13. Antil SK, Antil P, Singh S, Kumar A, Pruncu CI. Artificial neural network and response surface methodology based analysis on solid particle erosion P. Pattnaik, A. Sharma, M. Choudhary et al. Materials Today: Proceedings 44 (2021) 4703–47084707 behavior of polymer matrix composites. Materials (Basel), 2020, ;13, https://doi.org/10.3390/ma13061381.

14. V. Infante, JFA Madeira, R.B. Ruben, F. Moleiro, S.T. de Freitas, Characterization and optimization of hybrid carbon–glass epoxy composites under combined

loading, J. Compos. Mater.2019, 53 (18) 2593–2605, https://doi.org/10.1177/ 0021998319834673.

15. P.R. Pati, Prediction and wear performance of red brick dust filled glass–epoxy composites using neural networks, Int. J. Plast. Technol.2019, 23 (2) 253–260, https://doi.org/10.1007/s12588-019-09257-0.

16. P. Antil, S. Singh, A. Manna, Analysis on effect of electroless coated SiC p on mechanical properties of polymer matrix composites, Part. Sci. Technol. 2019, 37 (7) 791–798, https://doi.org/10.1080/02726351.2018.1444691.

17. Baseer, Ahmed Abdul, D. V. Ravi Shankar, and M. Manzoor Hussain. "Interfacial and tensile properties of hybrid FRP composites using dnn structure with optimization model." *Surface Review and Letters* 27.02 (2020): 1950099.https://doi.org/10.1142/S0218625X19500999.

18. Volovici, Victor, et al. "Steps to avoid overuse and misuse of machine learning in clinical research." *Nature Medicine, 2022,* 28.10 : 1996-1999, https://doi.org/10.1038/s41591-022-01961-6.

19. Verhelst, Hugo M., A. W. Stannat, and Giulio Mecacci. "Machine learning against terrorism: how big data collection and analysis influences the privacy-security dilemma." *Science and engineering ethics, 2020,* 26: 2975-2984, https://doi.org/10.1007/s11948-020-00254-w.

20. Gupta, Shivani, and Atul Gupta. "Dealing with noise problem in machine learning data-sets: A systematic review." *Procedia Computer Science, 2019,* 161 : 466-474, https://doi.org/10.1016/j.procs.2019.11.146.

21. Sarhan, Mhd Hasan, et al. "Machine learning techniques for ophthalmic data processing: a review." *IEEE Journal of Biomedical and Health Informatics, 2020,* 24.12: 3338-3350, https://doi.org/10.1109/JBHI.2020.3012134.

22. Ghosh, Pronab, et al. "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques." *IEEE Access* 9, 2021, 19304-19326, https://doi.org/10.1109/ACCESS.2021.3053759.

23. Yang, Li, and Abdallah Shami. "On hyperparameter optimization of machine learning algorithms: Theory and practice." *Neurocomputing,2020,* 415 : 295-316. https://doi.org/10.1016/j.neucom.2020.07.061.

24. Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science, 2021,* 2.3 : 160, https://doi.org/10.1007/s42979-021-00592-x.

25. Crawford, Kate, and Trevor Paglen. "Excavating AI: The politics of images in machine learning training sets." *Ai & Society, 2021,* 36.4 : 1105-1116, https://doi.org/10.1007/s00146-021-01162-8.

26. Chicco, Davide, and Giuseppe Jurman. "The ABC recommendations for validation of supervised machine learning results in biomedical sciences." *Frontiers in big Data* 5 : 2022, 979465, https://doi.org/10.3389/fdata.2022.979465.

27. Gui, Jie, et al. "A Survey on Self-supervised Learning: Algorithms, Applications, and Future Trends." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024,* https://doi.org/10.1109/TPAMI.2024.3415112.

28. Watson, David S. "On the philosophy of unsupervised learning." *Philosophy & Technology, 2023, 36.2 , 28.* https://doi.org/10.1007/s13347-023-00635-6.

29. Alimi, Oyeniyi Akeem, Khmaies Ouahada, and Adnan M. Abu-Mahfouz. "A review of machine learning approaches to power system security and stability." *IEEE Access* 8: 2020, 113512-113531, https://doi.org/10.1109/ACCESS.2020.3003568.

30. Hegde, Jeevith, and Børge Rokseth. "Applications of machine learning methods for engineering risk assessment–A review." *Safety science* 122, 2020, : 104492, https://doi.org/10.1016/j.ssci.2019.09.015.

31. Maulud, Dastan, and Adnan M. Abdulazeez. "A review on linear regression comprehensive in machine learning." *Journal of Applied Science and Technology Trends, 2020,* 1.4: 140-147, https://doi.org/10.38094/jastt1457.

32. Moodi, Yaser, Mohammad Ghasemi, and Seyed Roohollah Mousavi. "Estimating the compressive strength of rectangular fiber reinforced polymer–confined columns using multilayer perceptron, radial basis function, and support vector regression methods." *Journal of Reinforced Plastics and Composites* 41.3-4 : 2022, 130-146, https://doi.org/10.1177/07316844211050168.

33. Sujeeun, Lakshmi Yaneesha, et al. "Predictive modeling as a tool to assess polymer–polymer and polymer–drug interactions for tissue engineering applications." *Macromolecular Research, 2023,* 31.4 : 379-392, https://doi.org/10.1007/s13233-023-00155-4.

34. Wu, Jianxin, and Hao Yang. "Linear regression-based efficient SVM learning for large-scale classification." *IEEE transactions on neural networks and learning systems, 2015,* 26.10 : 2357-2369, https://doi.org/10.1109/TNNLS.2014.2382123.

35. Maulud, Dastan, and Adnan M. Abdulazeez. "A review on linear regression comprehensive in machine learning." *Journal of Applied Science and Technology Trends, 2020,* 1.4 : 140-147, https://doi.org/10.38094/jastt1457.

36. Balci, Onur, et al. "Prediction of CIELab data and wash fastness of nylon 6, 6 using artificial neural network and linear regression model." *Fibers and Polymers, 2008,* 9 : 217-224, https://doi.org/10.1007/s12221-008-0035-z.

37. QIU, Sai-bing, and T. A. N. G. Bo. "Application of mutiple linear regression analysis in polymer modified mortar quality control." 2nd International Conference on Electronic & Mechanical Engineering and Information Technology. Atlantis Press,2012, https://doi.org/10.2991/emeit.2012.245.

38. Ido, Ryota, et al. "A method for inferring polymers based on linear regression and integer programming." *arXiv preprint arXiv: 2021, 2109.02628* ,https://doi.org/10.48550/arXiv.2109.02628

39. Ghasemi, Hamid, and Hessam Yazdani. "Plastics and sustainability in the same breath: Machine learning-assisted optimization of coarse-grained models for polyvinyl chloride as a common polymer in the built environment." *Resources, Conservation and Recycling, 2022,* 186 : 106510, https://doi.org/10.1016/j.resconrec.2022.106510.

40. Ahmad, A. L., et al. "Prediction of plasticization pressure of polymeric membranes for CO2 removal from natural gas." Journal of Membrane Science, 2015, 480: 39-46, https://doi.org/10.1016/j.memsci.2015.01.039.

41. Yi, Yong, Liming Wang, and Zhengying Chen. "Adaptive global kernel interval SVR-based machine learning for accelerated dielectric constant prediction of polymer-based dielectric energy storage." *Renewable Energy, 2021,* 176: 81-88, https://doi.org/10.1016/j.renene.2021.05.045.

42. Arumugasamy, Senthil Kumar, et al. "Comparison between artificial neural networks and support vector machine modeling for polycaprolactone synthesis via enzyme catalyzed polymerization." *Process Integration and Optimization for Sustainability, 2021,* 5 : 599-607, https://doi.org/10.1007/s41660-021-00163-w.

43. Owolabi, Taoreed O., and Mohd Amiruddin Abd Rahman. "Modeling the optical properties of a polyvinyl alcohol-based composite using a particle swarm optimized support vector regression algorithm." *Polymers, 2021,* 13.16 : 2697, https://doi.org/10.3390/polym13162697.

44. Rana, Nadim, et al. "A Systematic Literature Review on Contemporary and Future trends in Virtual Machine Scheduling Techniques in Cloud and Multi-Access Computing." , 2023, https://doi.org/10.21203/rs.3.rs-2792348/v1.

45. Amor, Nesrine, Muhammad Tayyab Noman, and Michal Petru. "Classification of textile polymer composites: Recent trends and challenges." *Polymers, 2021,* 13.16 : 2592. https://doi.org/10.3390/polym13162592.

46. Koinig, G., et al. "Inline classification of polymer films using Machine learning methods." *Waste Management, 2024,* 174: 290-299. https://doi.org/10.1016/j.wasman.2023.11.028.

47. Yang, Zhen, et al. "Structure control classification and optimization model of hollow carbon nanosphere core polymer particle based on improved differential evolution support vector machine." *Applied Mathematical Modelling, 2013,* 37.12-13: 7442-7451, https://doi.org/10.1016/j.apm.2013.02.037.

48. Altarazi, Safwan, Rula Allaf, and Firas Alhindawi. "Machine learning models for predicting and classifying the tensile strength of polymeric films fabricated via different production processes." Materials., 12.9, 2019, 1475, https://doi.org/10.3390/ma12091475.

49. Zhang, Yiyi, et al. "Moisture prediction of transformer oil-immersed polymer insulation by applying a support vector machine combined with a genetic algorithm." Polymers 12.7 : 2020, 1579, https://doi.org/10.3390/polym12071579.

50. Lorenzo-Navarro, Javier, et al. "Performance evaluation of classical classifiers and deep learning approaches for polymers classification based on hyperspectral images." *Advances in Computational Intelligence: 16th International Work-Conference on Artificial Neural Networks, IWANN 2021, Virtual Event, June 16–18, 2021, Proceedings, Part II 16*. Springer International Publishing, 2021, https://doi.org/10.1007/978-3-030-85099-9_23.

51. Tian, Xin, Frederic Beén, and Patrick S. Bäuerlein. "Quantum cascade laser imaging (LDIR) and machine learning for the identification of environmentally exposed microplastics and polymers." Environmental Research 212: 2022, 113569. https://doi.org/10.1016/j.envres.2022.113569.

52. S. Panchal, A. Phadke, U. Gopinathan and S. Datar, "Development of a Polymer Modified Quartz Tuning Fork (QTF) Sensor Array-Based Volatile Organic Compound (VOC) Classifier," in *IEEE Sensors Journal*, 2021, vol. 21, no. 18, pp. 20870-20877, 15 Sept.15, https://doi.org/10.1109/JSEN.2021.3099537.

53. Dilbas, Hasan, et al. "Mechanical performance improvement of super absorbent polymer-modified concrete." MethodsX 10 : 2023, 102151, https://doi.org/10.1016/j.mex.2023.102151.

54. Das, Abhik. "Logistic regression." *Encyclopedia of Quality of Life and Well-Being Research*. Cham: Springer International Publishing. 1-2, 2021, https://doi.org/10.1007/978-3-319-69909-7_1689-2.

55. Boateng, Ernest Yeboah, and Daniel A. Abaye. "A review of the logistic regression model with emphasis on medical research." *Journal of data analysis and information processing* 7.4, 2019, 190-207, https://doi.org/10.4236/jdaip.2019.74012.

56. Zhang, J.Z. Development of an in-process Pokayoke system utilizing accelerometer and logistic regression modeling for monitoring injection molding flash. Int J Adv Manuf Technol 71, 2014, 1793–1800, https://doi.org/10.1007/s00170-013-5604-7.

57. Idolor, Ogheneovo, et al. "Nondestructive examination of polymer composites by analysis of polymer-water interactions and damage-dependent hysteresis." Composite Structures 287 : 2022, 115377, https://doi.org/10.1016/j.compstruct.2022.115377.

58. Mairpady, Anusha, Abdel-Hamid I. Mourad, and Mohammad Sayem Mozumder. "Accelerated Discovery of the Polymer Blends for Cartilage Repair through Data-Mining Tools and Machine-Learning Algorithm." Polymers 14.9 : 2022, 1802, https://doi.org/10.3390/polym14091802.

59. Xu, Xin, et al. "Recognition of polymer configurations by unsupervised learning." *Physical Review E* 99.4: 2019, 043307. https://doi.org/10.1103/PhysRevE.99.043307.

60. Verma, Prabha, and R. D. S. Yadava. "Polymer selection for SAW sensor array based electronic noses by fuzzy c-means clustering of partition coefficients: Model studaies on detection of freshness and spoilage of milk and fish." *Sensors and Actuators B: Chemical* 209 : 2015, 751-769, https://doi.org/10.1016/j.snb.2014.11.149.

61. Chen, Zhudan, et al. "Unsupervised machine learning methods for polymer nanocomposites data via molecular dynamics simulation." *Molecular Simulation* , 2020, 46.18: 1509-1521, https://doi.org/10.1080/08927022.2020.1851028.

62. Tang, Yun, et al. "Industrial polymers classification using laser-induced breakdown spectroscopy combined with self-organizing maps and K-means algorithm." *Optik* 165 : 2018, 179-185, https://doi.org/10.1016/j.ijleo.2018.03.121.

63. Pezer, Danijela. "Application of Cluster Analysis for Polymer Classification According to Mechanical Properties." , https://orcid.org/0000-0002-0787-948X, (2021).

64. Kurita, Hiroki, et al. "k-Means Clustering for Prediction of Tensile Properties in Carbon Fiber-Reinforced Polymer Composites." *Advanced Engineering Materials,* 2022, 24.5: 2101072, https://doi.org/10.1002/adem.202101072.

65. Karamizadeh, Sasan, et al. "An overview of principal component analysis." *Journal of Signal and Information Processing* 4.3B : 2013, 173, https://doi.org/10.4236/jsip.2013.43B031.

66. Hiles, Melanie, Michael Grossutti, and John R. Dutcher. "Classifying formulations of crosslinked polyethylene pipe by applying machine-learning concepts to infrared spectra." Journal of Polymer Science Part B: Polymer Physics, 2019, 57.18 : 1255-1262, https://doi.org/10.1002/polb.24837.

67. Lightstone, Jordan P., et al. "Refractive index prediction models for polymers using machine learning." Journal of Applied Physics, 2020, 127.21.5.0008026/10.1063 https://doi.org/10.1063/5.0008026.

68. Bhowmik, Rahul, et al. "Prediction of the specific heat of polymers from experimental data and machine learning methods." Polymer 220: 2021, 123558. https://doi.org/10.1016/j.polymer.2021.123558.

69. Gupta, Praveen, Linda S. Schadler, and Ravishankar Sundararaman. "Dielectric properties of polymer nanocomposite interphases from electrostatic force microscopy using machine learning." Materials Characterization 173: 2021, 110909, https://doi.org/10.1021/acsaelm.2c01331.

70. I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," in *IEEE Access*, vol. 10, pp. 2022, 99129-99149, https://doi.org/10.1109/ACCESS.2022.3207287.

71. Liu C, Liao J, Zheng Y, Chen Y, Liu H, Shi X. Random forest algorithm-enhanced dual-emission molecularly imprinted fluorescence sensing method for rapid detection of pretilachlor in fish and water samples. J Hazard Mater. Oct 5;439:129591, Epub 2022 Jul 14. PMID: 35853336. https://doi.org/10.1016/j.jhazmat.2022.129591.

72. Wang, Wenyang, and Dongchu Sun. "The improved AdaBoost algorithms for imbalanced data classification." *Information Sciences, 2021, 563, 358-374*. https://doi.org/10.1016/j.ins.2021.03.042.

73. Wang, Qichen, et al. "Application of soft computing techniques to predict the strength of geopolymer composites." *Polymers, 2022,* 14.6 : 1074. https://doi.org/10.3390/polym14061074.

74. Amin, Muhammad Nasir, et al. "Use of artificial intelligence for predicting parameters of sustainable concrete and raw ingredient effects and interactions." *Materials, 2022,* 15.15: 5207, https://doi.org/10.3390/ma15155207.

75. Anjum, Madiha, et al. "Application of ensemble machine learning methods to estimate the compressive strength of fiber-reinforced nano-silica modified concrete." *Polymers, 2022,* 14.18 3906, https://doi.org/10.3390/polym14183906.

76. Ledwani, D., I. Thakur, and V. Bhatnagar. "Comparative Analysis of Prediction Models for Melt Flow Rate of C2 and C3 Polymers Synthesized using Nanocatalysts." *NanoWorld J* 8.S1: 2022, S123-S127, https://doi.org/10.17756/nwj.2022-s1-022.

77. Sheng, Wenjuan, Yutao Liu, and Dirk Söffker. "A novel adaptive boosting algorithm with distance-based weighted least square support vector machine and filter factor for carbon fiber reinforced polymer multi-damage classification." Structural Health Monitoring 22.2 (2022): 1273-1289, https://doi.org/10.1177/14759217221098173.

78. Yarahmadi, Bita, Seyed Majid Hashemianzadeh, and Seyed Mohammad-Reza Milani Hosseini. "A new approach to prediction riboflavin absorbance using imprinted polymer and ensemble machine learning algorithms." *Heliyon, 2023,* 9.7.

79. Ahmed, Shams Forruque, et al. "Deep learning modelling techniques: current progress, applications, advantages, and challenges." *Artificial Intelligence Review* , 2023, *56.11* , *13521-13617.* https://doi.org/10.1007/s10462-023-10466-8.

80. Mengshan, Li, et al. "Solubility prediction of gases in polymers based on an artificial neural network: a review." *RSC advances* 7.56 : 35274-35282, https://doi.org/10.1039/C7RA04200K, (2017).

81. Sharma, Aanchna, and Vinod Kushvaha. "Predictive modelling of fracture behaviour in silica-filled polymer composite subjected to impact with varying loading rates using artificial neural network." Engineering Fracture Mechanics 239 : 107328. https://doi.org/10.1016/j.engfracmech.2020.107328, (2020).

82. Wang, Qiang, et al. "Superhydrophobic polymer topography design assisted by machine learning algorithms." *ACS Applied Materials & Interfaces* 13.25 : 30155-30164, https://doi.org/10.1021/acsami.1c04473, (2021).

83. Haddad, Rami, and Madeleine Haddad. "Predicting fiber-reinforced polymer–concrete bond strength using artificial neural networks: A comparative analysis study." Structural Concrete 22.1 : 38-49, https://doi.org/10.1002/suco.201900298, (2021).

84. Amor, Nesrine, Muhammad Tayyab Noman, and Michal Petru. "Classification of textile polymer composites: recent trends and challenges." *Polymers* 13.16 : 2592. https://doi.org/10.3390/polym13162592, (2021).

85. Koushik, Jayanth. "Understanding convolutional neural networks." *arXiv preprint arXiv:1605.09081* , https://doi.org/10.48550/arXiv.1605.09081, (2016).

86. Zeng, Minggang, et al. "Graph convolutional neural networks for polymers property prediction." *arXiv preprint arXiv:1811.06231* , https://doi.org/10.48550/arXiv.1811.06231, (2018).

87. Ma, Liyong, Wei Xie, and Yong Zhang. "Blister defect detection based on convolutional neural network for polymer lithium-ion battery." Applied Sciences 9.6: 1085, https://doi.org/10.3390/app9061085, (2019).

88. Miccio, Luis A., and Gustavo A. Schwartz. "From chemical structure to quantitative polymer properties prediction through convolutional neural networks." *Polymer* 193 : 122341, https://doi.org/10.1016/j.polymer.2020.122341, (2020).

89. Taş, Göksu, Ali Uysal, and Cafer Bal. "A New Lithium Polymer Battery Dataset with Different Discharge Levels: SOC Estimation of Lithium Polymer Batteries with Different Convolutional Neural Network Models." *Arabian Journal for Science and Engineering* 48.5: 6873-6888, https://doi.org/10.1007/s13369-022-07586-8, (2023).

90. Jia, Z., et al. (2025). Machine learning helps chemists develop tougher plastics. *MIT News*.

91. Malashin, I., et al. (2025). Boosting-Based Machine Learning Applications in Polymer Science. *Polymers*, 17(4), 499. **https://doi.org/10.3390/polym17040499**

92. Kumar, J. N, Li Qianxiao, Jun Ye. "Challenges and opportunities of polymer design with machine learning and high throughput experimentation" *MRS Communications* 9(2), 537-544. https://doi.org/10.1557/mrc.2019.54 , (2019).

93. Zhao, Junnan, Aaron C. Tan, and Peter F. Green. "Thermally induced chain orientation for improved thermal conductivity of P (VDF-TrFE) thin films." *Journal of Materials Chemistry C* 5.41, 10834-10838. https://doi.org/10.1039/C7TC03240D, (2017).

94. Miccio L. A, Schwartz G. A. "From chemical structure to quantitative polymer properties prediction through convolutional neural networks". *Polymer.* 193, 122341. https://doi.org/10.1016/j.polymer.2020.122341, (2020).

95. Chen, Hongyu, et al. "Thermal conductivity of polymer-based composites: Fundamentals and applications." *Progress in Polymer Science* 59, 41-85. https://doi.org/10.1016/j.progpolymsci.2016.03.001, (2016).

96. González, Carlos, and Joaquín Fernández-León. "A machine learning model to detect flow disturbances during manufacturing of composites by liquid moulding." *Journal of Composites Science* 4.2 , 71. https://doi.org/10.3390/jcs4020071, (2020).

97. Cassola, Stefano, et al. "Machine learning for polymer composites process simulation–a review." *Composites Part B: Engineering 246, 110208.* https://doi.org/10.1016/j.compositesb.2022.110208, (2022).

98. Xu, Qisong, et al. "Synergizing machine learning, molecular simulation and experiment to develop polymer membranes for solvent recovery." *Journal of Membrane Science* 678 , 121678. https://doi.org/10.1016/j.memsci.2023.121678, (2023).

99. Wheatle, Bill K., et al. "Design of polymer blend electrolytes through a machine learning approach." *Macromolecules* 53.21, 9449-9459. https://doi.org/10.1021/acs.macromol.0c01547, (2020).

100. Dhanorkar, Ritesh J., Subhra Mohanty, and Virendra Kumar Gupta. "Synthesis of functionalized styrene butadiene rubber and its applications in SBR–silica composites for high performance tire applications." *Industrial & Engineering Chemistry Research* 60.12 , 4517-4535. https://doi.org/10.1021/acs.iecr.1c00013, (2021).

101. Hejna, Aleksander, et al. "Waste tire rubber as low-cost and environmentally-friendly modifier in thermoset polymers–A review." *Waste Management* 108, 106-118. https://doi.org/10.1016/j.wasman.2020.04.032, (2020).

102. Verma, Akarsh, et al. "Processing and characterization analysis of pyrolyzed oil rubber (from waste tires)-epoxy polymer blend composite for lightweight structures and coatings applications." *Polymer Engineering & Science* 59.10, 2041-2051. https://doi.org/10.1002/pen.25204, (2019).

103. Pang, Song, Jinlian Luo, and Youping Wu. "Properties prediction and design of tire tread composites using machine learning." *Macromolecular Theory and Simulations* 29.3 , 1900063.https://doi.org/10.1002/mats.201900063, (2020).

104. Hu, Yusha, et al. "Fuel composition forecasting for waste tires pyrolysis process based on machine learning methods." *Fuel* 362, 130789. https://doi.org/10.1016/j.fuel.2023.130789, (2024).

105. Correia, Sivaldo Leite, Denilso Palaoro, and Ana Maria Segadães. "Property optimisation of epdm rubber composites using mathematical and statistical strategies." *Advances in Materials Science and Engineering* 2017.1,, 2730830. https://doi.org/10.1155/2017/2730830, (2017).

106. Ghaffarian, N., and M. Hamedi. "Optimization of rubber compound design process using artificial neural network and genetic algorithm." *International Journal of Engineering 33.11, 2319-2326.* https://doi.org/10.5829/ije.2020.33.11b.22, (2020).

107. Saran, O. Sai, et al. "3D printing of composite materials: A short review." *Materials Today: Proceedings* 64, 615-619. https://doi.org/10.1016/j.matpr.2022.05.144, (2022).

108. Dananjaya, S. A. V., et al. "3D printing of biodegradable polymers and their composites–Current state-of-the-art, properties, applications, and machine learning for potential future applications." *Progress in Materials Science*, 101336. https://doi.org/10.1016/j.pmatsci.2024.101336, (2024).

109. Karkaria, V., et al. (2024). A Machine Learning–Based Tire Life Prediction Framework for Increasing Life of Commercial Vehicle Tires. *ASME Journal of Computing and Information Science in Engineering*. https://doi.org/10.1115/1.4063761

110. Jia, Z., et al. (2025). Machine learning helps reveal key factors affecting tire wear. *Science of The Total Environment*. https://doi.org/10.1016/j.envint.2024.109224

111. Tong, Z., et al. (2025). Machine learning-driven intelligent tire wear detection system. *Computers in Industry*. https://doi.org/10.1016/j.measurement.2024.115848

112. Ferdousi, Sanjida, et al. "Investigation of 3D printed lightweight hybrid composites via theoretical modeling and machine learning." *Composites Part B: Engineering* 265, 110958. https://doi.org/10.1016/j.compositesb.2023.110958 ,(2023).

113. Dahiya, Neelam, Sheifali Gupta, and Sartajvir Singh. "A review paper on machine learning applications, advantages, and techniques." *ECS Transactions* 107.1: 6137. https://doi.org/10.1149/10701.6137ecst, (2022).

114. Tan, Yu Guang, et al. "Incorporating artificial intelligence in urology: Supervised machine learning algorithms demonstrate comparative advantage over nomograms in predicting biochemical recurrence after prostatectomy." *The Prostate* 82.3: 298-305. https://doi.org/10.1002/pros.24272, (2022).

115. Hassan, Ch Anwar Ul, Muhammad Sufyan Khan, and Munam Ali Shah. "Comparison of machine learning algorithms in data classification." *24th International Conference on Automation and Computing (ICAC)*. IEEE. 10.23919/IConAC.2018.8748995 (2018).

116. Jha, Khushi Kumari, et al. "A brief comparison on machine learning algorithms based on various applications: a comprehensive survey." *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE. http://10.1109/CSITSS54238.2021.9683524, (2021).