SUPERVISED LEARNING MODELS FOR THE PREDICTION OF MATERIAL PROPERTIES

Abstract

Machine learning (ML) techniques play a major role in engineering world. In this sequence the manufacturing industries also utilize the ML techniques for the various applications. Among them material properties prediction or forecasting is a noticeable process of manufactures using ML techniques. The ML techniques are broadly categorized into three types such as supervised; semi supervised and unsupervised learning techniques. The learning approach can be preferred based on the problem to solve using ML technique. In this chapter, the supervised learning for the prediction of material properties is presented. Initially the properties of materials and the necessity of ML technique for the prediction of material properties is described. Then four different supervised learning such as Random Forest (RF), Naive Bayesian (NB), Support Vector Machine (SVM), and Artificial Neural Network (ANN) are described for the prediction of material properties. Finally, the performance of these four techniques is evaluated based on accuracy. The performance analysis shows that the ANN with accuracy of 98% provides better than other techniques.

Keywords: supervised learning, material properties, prediction, application of machine learning

Authors

Arivarasi A
Micro Nano Devices Research group
School of Electronics (SENSE)
Vellore Institute of Technology
Chennai, India

Gayathri S
Micro Nano Devices Research group
School of Electronics (SENSE)
Vellore Institute of Technology
Chennai, India

Sathya Sree J
Micro Nano Devices Research group
School of Electronics (SENSE)
Vellore Institute of Technology
Chennai, India

I. INTRODUCTION

The field of material properties depends on experimental and simulation-based models as tools for material characterization. Material properties, which including behaviour and structure, are essential to a material's application potential. In recent times, data generated by such simulations and experiments has opened up a different possibilities for data-driven methodologies to be used. In recent years, In addition to the new trial-and-error approach or the physical metallurgy approach, Machine learning (ML) methods for material design and property prediction have generated a lot of attention [1].

Experiments have been conducted since the stone and copper periods, it is not until the 16th century that scientists start using equations to describe physical relationships. Therefore, analytical equations become a great component in theoretical physics, complementing which are able to empirical and experimental sciences. The third paradigm developed at the start of simulations and computational materials science. Computer simulations and experiments were available under this framework, with the generated data being processed evaluated in the same way as observable data. It had to be recognized that many material properties are committed to multiple multilevel, sophisticated theoretical concepts that does not expressed by a closed mathematical form. Though, it is not only advantageous to have a great amount of data, dealing with large amounts of data can also be difficult. In machine learning, several techniques were developed to predict the material property. Here, four techniques are used to predict the material.

What are Material Properties?

A materials property is defined as some materials of intensive property, i.e., a physical property that is independent of the material's quantity. These quantitative qualities can be used as a metric to compare the advantages of various materials, which can help in material selection. A property could be a function of one or more independent variables or a constant, such as temperature. Anisotropy is a condition in which a substance's properties change to some extent depending on the direction in which it is measured [2]. Materials qualities transmit to various physical phenomena frequently behave linearly within specific working range, so more information is required. The property is shown using differential constitutive equations that are represented as linear functions, they can be substantially simplified. Material properties are the chemical, physical, or mechanical characteristics of a specific product that determine its usability and manufacturing feasibility [3]. This would suggest that a product's material qualities would determine the product's overall capabilities. Some of the material properties list are briefly explained in the below section.

1. **Strength:** the ability to withstand forces without breaking
2. **Hardness:** ability to withstand scratching and abrasion
3. **Toughness:** hard to break or snap
4. **Elasticity :** can stretch or bend then return to its original shape
5. **Malleability:** can be bent or shaped
6. **Ductility :** can be drawn (pulled) into long wires
7. **Conductivity:** ability to let heat or electricity flow through
8. **Fusibility :** have a low melting point
9. **Density :** mass per unit volume
10. **Absorbency:** good at soaking up moisture

II. PURPOSE OF PREDICTING MATERIAL PROPERTIES

In forming materials, considerate the material's properties can help to better predict the manufacturing outcome. A material profile will help detect material variations from suppliers for stamping and forming processes, as well as for diverse products [4]. The number of studies systematically investigating varied families of materials with the purpose of identifying previously unknown features in current materials or developing new materials with desired properties is continually increasing. Experimental structure databases, particularly crystalline structure databases, are continuously expanding, with the addition of larger and larger databases of physical and chemical attributes [5]. High-throughput research and combinatorial materials synthesis are made possible by robotics and artificial intelligence, allowing for faster reactions and analysis.

In terms of computation, Scientists can already forecast the structure and properties of complex materials before they are made using molecular simulations, which have increased in scale. To support in the search for new materials, high-throughput screening techniques can be combined with compound property prediction. Yet, despite increases in computational capacity, whether classical or quantum, computational methods, are still time consuming and can only scratch, surface attributes of all potential chemical compositions and crystal structures. Machine learning (ML), is a type of Artificial Intelligence (AI) that has witnessed great improvements in recent years and widespread use across many fields of research, is becoming increasingly popular in computational chemistry and materials science, to attain this objective of systematic chemical space exploration and to make use of the large-scale databases of structures and attributes that are now available. As in previous section the introduction of material property and some list is explained briefly. Following that the supervised learning models and types will be discussed briefly.

III. SUPERVISED LEARNING MODELS

Supervised learning is a method by use labeled training data to train a function that generalize for new examples. In supervised learning, construct a function with both input and output data from labelled training data [6]. A model in machine learning, the learning process separated into two steps: training and testing. In training process, the learning model is developed, samples from training data are used as input, with structures learned by the learning algorithm or learner. In testing process, the performance engine is used by the learning model to generate predictions for experiments are production data. Then, the output of a learning model is identified data; it is the categorized data or final prediction. In this architecture diagram is represents supervised learning process and it shows the figure 1.

The method of supervised learning is maximum simple technique in the classification issue; the aim is generally to train the machine a classification method that created the data.

In maximum cases, supervised learning leaves the possibility of an input, including such as one with a known projected outcome, unknown. This procedure generates a dataset with labels and features. The main goal is to develop an estimator that can estimate the label of an object based on a collection of attributes [7]. Then, as inputs, the learning algorithm is given a collection of features as well as the correct outputs, and it learns by detecting faults by comparing its true output to the corrected outputs. It then performs the appropriate model

variations. It is not essential to develop a model as long as the inputs are available; but, it's impossible to deduce anything about the outputs if any of the input values are absent.
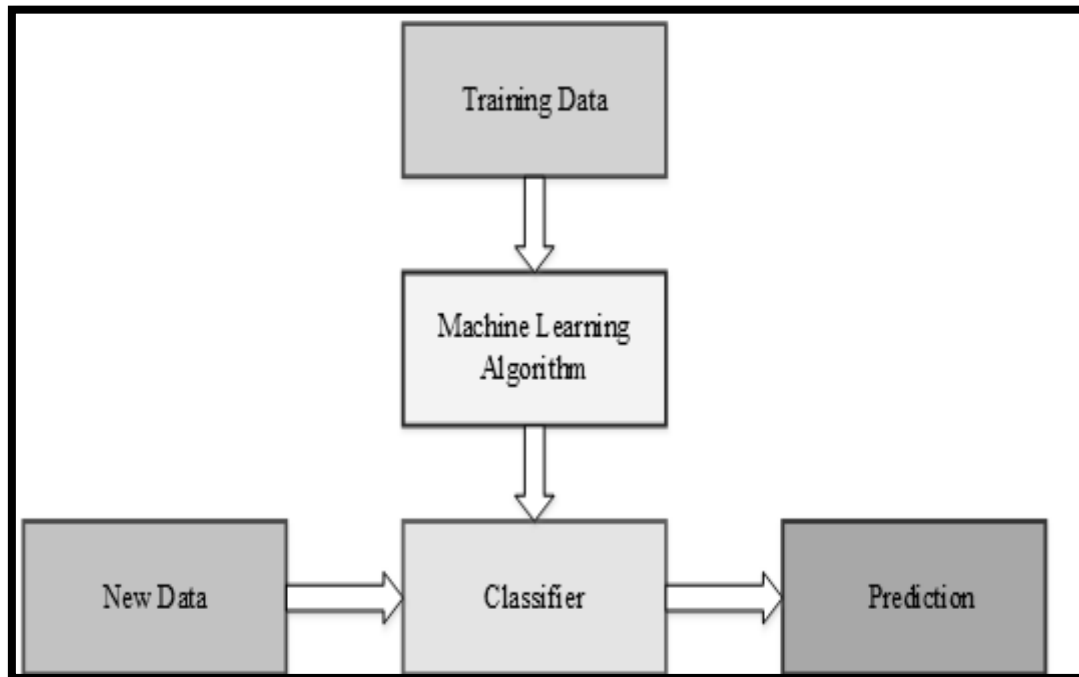

Figure 1: Supervised Learning Process

Supervised learning is the most common method for training neutral networks and decision trees. The dependent data collected through the pre-determined classification was compared. Similarly, this learning strategy is most commonly employed in applications, because previous data predicts likely feature behaviors. This learning has a number of practical applications, including one that guesses the species of iris based on a collection of flower measurement [8]. The supervised learning responsibilities are divided into two groups, as explained previously: regression and classification. The label in classification is disconnected, but the label in regression is continuous.

As shown on Figure 2, the method distinguishes between observed data X and training data, which is often structured data fed into the model during training. In this procedure, the predictive model is built using the supervised learning technique. Once trained, the fitted model would attempt to predict the most likely labels for a new set of samples X in the testing set. The type of target y can be used to classify supervised learning methods:

1. The task of predict y is called classification if y is a fixed collection of categorical outcomes (integers).
2. Regression assignment to predict y, when y has floating point values.
3. In previous section, supervised learning process are explained in detail and types of supervised learning are briefly discussed in upcoming section.
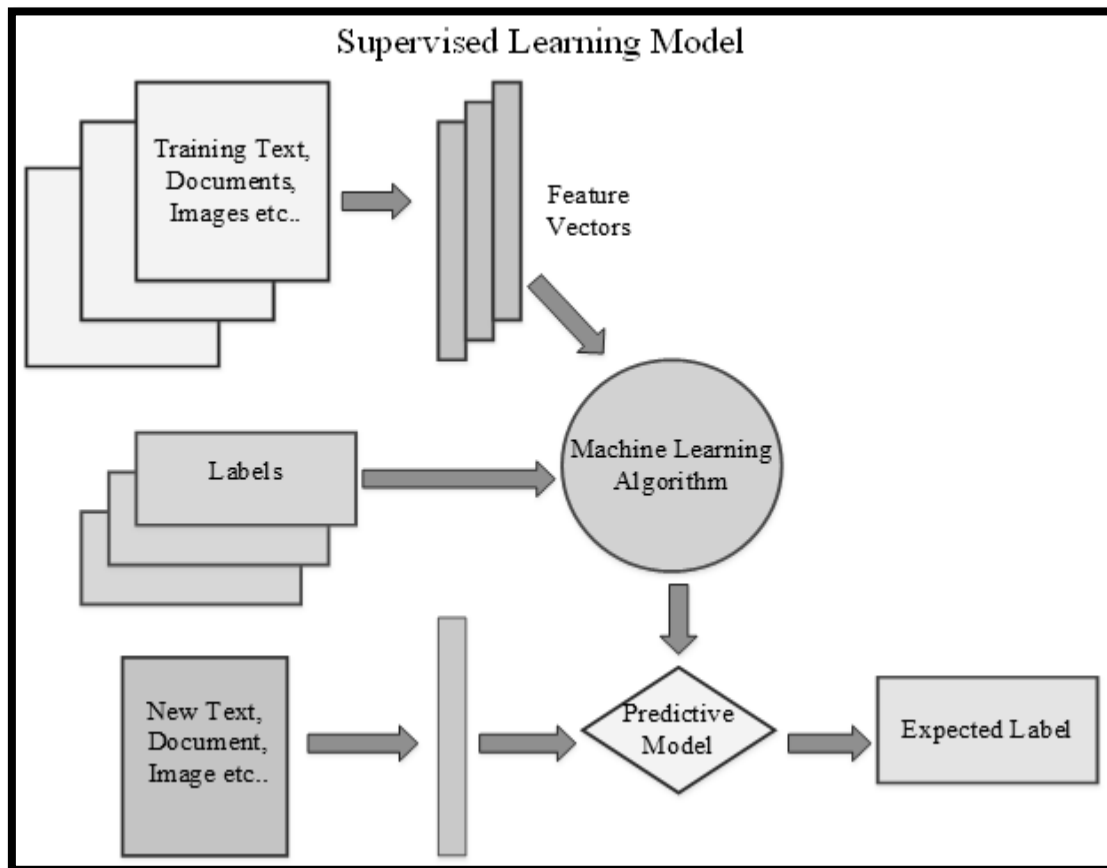
Figure 2: Supervised Learning Model

IV. TYPES OF SUPERVISED LEARNING MODELS

Supervised learning is a type of machine learning in which machines are trained to predict outcomes using well-labeled training data [9]. Data that has already been classed with the correct output is referred to as "labelled data". Some types of supervised learning models and architecture are briefly conversed in below:

1. **Naive Bayesian (NB) Networks:** The most basic Bayesian networks are directed acyclic graphs with only one parent (representing the unseen node) and many children (representing observed nodes), with a strong assumption of child node individual freedom in their parent's environment. Therefore, the Naive Bayes model of independence is based on approximating. When compared to more advance learning algorithms, Bayes classifiers are frequently less accurate (such as ANNs). But, a large-scale comparison of the naive Bayes classifier with state-of-the-art decision tree induction approaches, it was observed that instance-based learning is sometimes stronger to other learning methodologies, even on datasets with significant feature dependent variables [10]. Attribute independence problem in Bayes classifiers was solved using Averaged One-Dependence Estimators.

   Consider the $P(x_1, x_2)$ generic probability distribution. Get this equation by applying the Bayes rule while preserving generality.

$$P(x_1, x_2) = P(x_1|x_2)P(x_2) \tag{1}$$

Get, following equation if there is another class variable c:

$$P(x_1, x_2|c) = P(x_1|x_2, c)P(x_2|c) \tag{2}$$

If scenario is generalized from two variables to a set of variables with conditional independence $x_1, \ldots x_N$ conditional on another variable c, get the following:

$$P(x|c) = \prod_{i=1}^{N} P(x_i|c) \tag{3}$$

2. **Decision tress:** Decision Trees are a type of Supervised Machine Learning in which data is continually separated based on a parameter. The tree can be described by two multiple individuals: leaves and decision nodes. A classifier is represented as a decision tree, which is a recursive split of the instance space. Root tree, with no incoming edges and a root node that is distributed, is made up of nodes that make up the decision tree. There is exactly one incoming edge for each other node. An internal node, often known as a test node, has outbound edges. The remaining nodes are referred to as leaves. Based on the discrete functions of the input values, in a decision tree, the instance space is divided into two or more sub-spaces by each test node. Each analysis identifies a single attribute in simplest scenario, and the instance space is partitioned into portions based on the property's value [11]. The condition in the case of numeric properties refers to a range. Each leaf is given the class that best corresponds to the goal value. A probability vector representing the probability of the target characteristic having a specific value should be present in the leaf. The situations are grouped based on the results of the tests conducted along the way, which are investigated from the tree's root to its leaf.
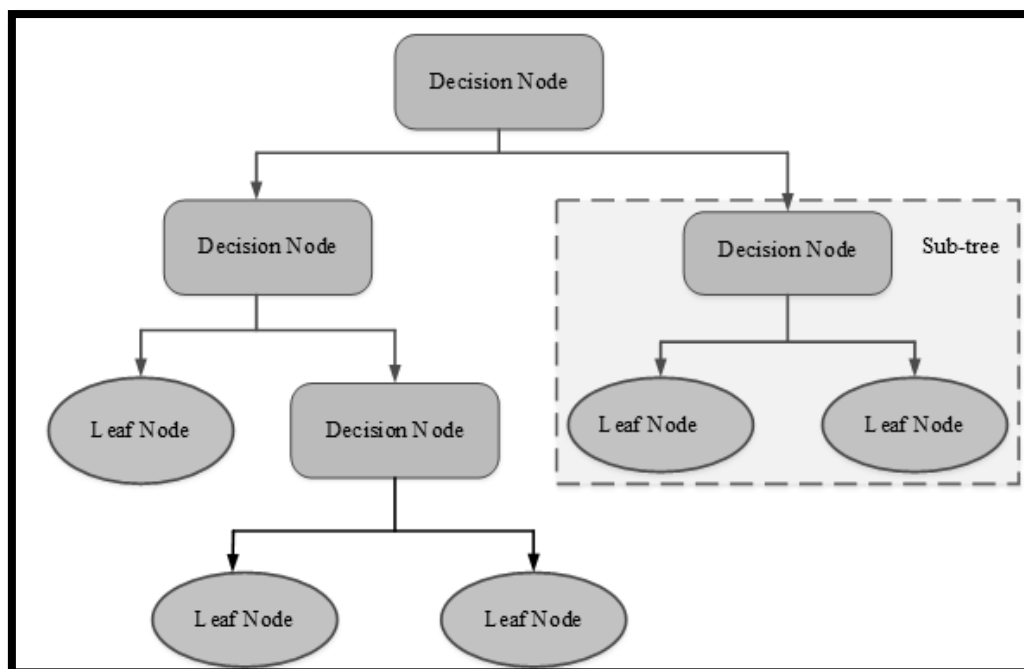

Figure 3: Decision Tress for Machine Learning

In machine learning, use past data to predict a future state. The classification technique data are desired attribute is defined as supervised learning. Many algorithms exist to help with this type of learning and decision tree is the one of the example. A decision tree is a directed graph, the nodes represent attribute tests, branches represent test outcomes, and leaves represent class labels.

3. **Support vector machine:** SVMs were designed to solve classification and regression analysis difficulties in the beginning. In supervised learning SVM is one of the technique that has been developed to classify data from different sources. These are applied to both non-linear and linear data classification tasks and employed for two-class classification problems. In a high-dimensional space, SVM generates a hyperplane or a group of hyperplanes, with the greatest hyperplane splitting data into distinct groups with the greatest separation. To anticipate the margins, a non-linear classifier. The fundamental goal of these kernel functions (linear, sigmoid, polynomial and radial basis) maximize hyperplane margins. Newly, researchers have developed a number of very promising applications as a result increased interest in SVMs. In a long time, SVM has been used in image processing and pattern recognition [12]. The diagram represents the support vector machine is shown in figure 4.
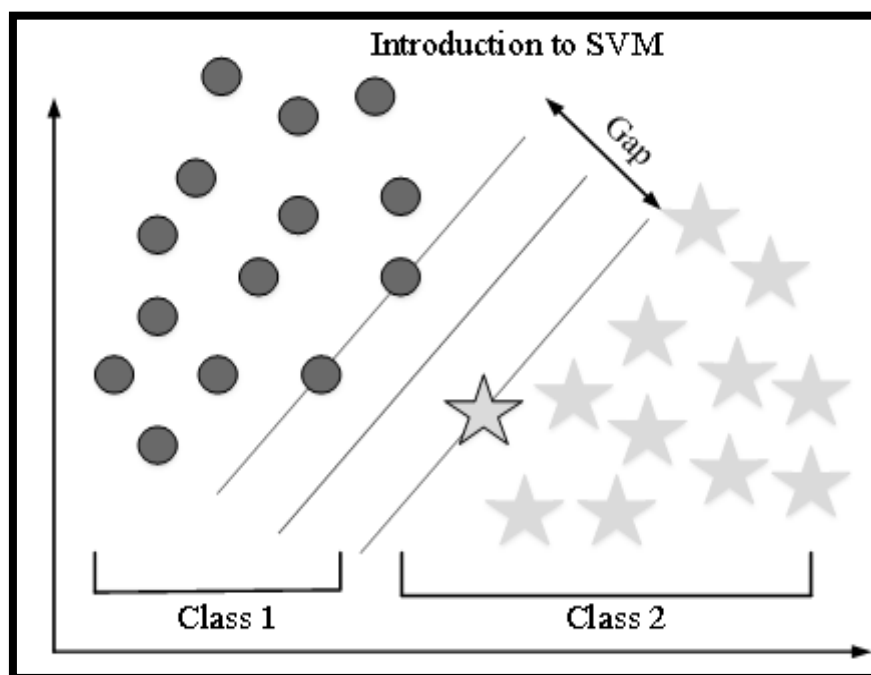

Figure 4: Diagram of Support vector machine

Change the data into a multi-dimensional feature, the Support Vector Machine (SVM) is a popular method in data analysis, pattern recognition, classification and regression. Considering a data series for regression, $(x_1, y_1), (x_2, y_2), . (x_N, y_N)$ the total number of data samples is represented by N. The regression function of SVM produced as [13]:

$$f(x) = w^T \Phi(X) + b \qquad (4)$$

Where w a vector in the feature is space F and $\Phi(X)$ is the feature map set the input data x to a vector in F. Important a loss function that is ε insensitive.

$$L_\varepsilon(x, y, f) = |y - f(x)|_\varepsilon = max(0, |y - f(x)| - \varepsilon) \tag{5}$$

The following equation can be used to calibrate the parameters w and b:

$$C \sum_{i=1}^{N} L_\varepsilon(x_i, y_i, f) + \frac{1}{2}\|w\|^2 \tag{6}$$

Where ε indicate maximum deviation, C represents for training drawback, which measures the model smoothness and trade-off between empirical risks.

4. **Random forest:** The random forest and bootstrapping selection methods are combined in random forest approach. Each tree in the bootstrapping method grows with a unique training sample drawn at random from the training database. Some observations may appear more than once in the bootstrap sample due to sampling with replacement, while others will be eliminated, known as out-of-bag (OOB) observations. The bootstrapping technique uses random feature selection to ignore any possible connection between the base trees. At each splitting node, it simply allows for a random selection of variables, rather than employing all of the explanatory variables [14]. To ensure single tree variability, RF typically employs two degrees of randomness: (a) a separate data set for training with the sample size is same, and (b) To separate each node, a distinct set of explanatory variables was used. Because only a subset of samples and explanatory variables were employed, the prediction errors caused by skewed samples and noisy data could be reduced. Intelligence, RF has a strong technical requirement, and data variation is selected. The prediction of RF mainly shows three characteristics: (a) correlation between base trees: mainly need decrease the correlation; (b) every tree has good effectiveness: Each base tree's performance need to increase; and (c) overall tress in the area: Due to computational efficiency considerations, the number must be large. More definitely, the three parameters are the total number of trees n (forest size), the number of splitting variables m, and the maximum tree depth d.
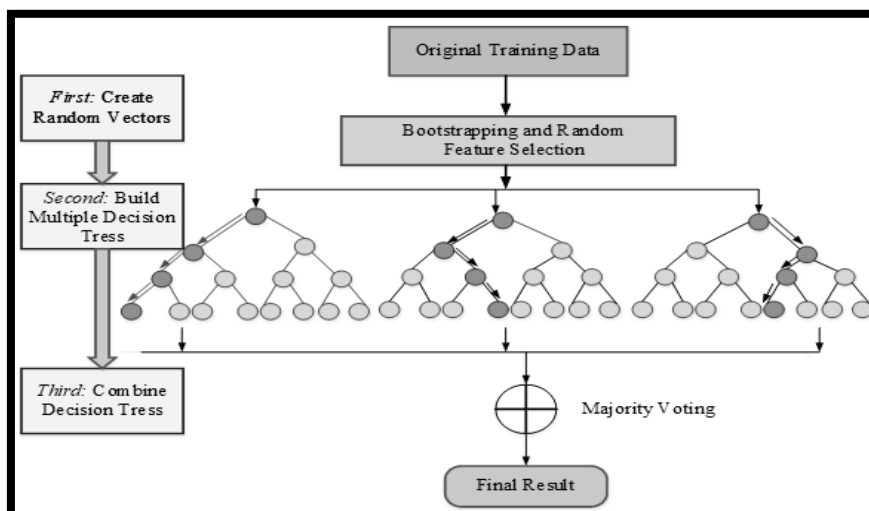

Figure 5: Architecture of Random Forest Method

In figure 5, explains the working process of random forest. For every tree, various bootstrapped and observations. The applicant splitting variables for each tree are chosen at random from a list of explanatory variables. Splitting continues until every individual tree reaches its maximum depth. The result is calculated using the majority voting approach once the base tree models have been estimated. The final anticipated outcome that ensemble predicts the most.

5. **Artificial neural network:** The ANN consists of the layer is $L + 1$ and it contains two types (i.e.) layer L is the output layer and layer 0 is the input layer. The hidden layers is $0 < l < L$. Activation function of hidden layer, such as hyperbolic tangents, rectified linear units, or sigmoids. Use sigmoids in the buried layers unless otherwise specified. At last, the output activation will be the linear activation. ANN defines mapping$\mathbb{R}^N \rightarrow \mathbb{R}^M$.
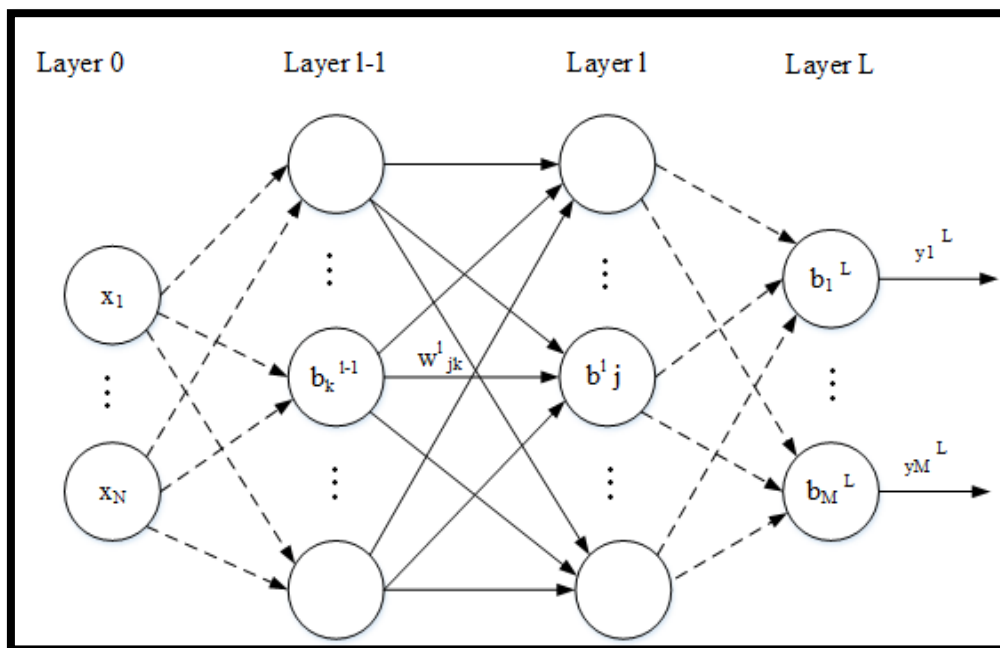

Figure 6: Architecture of a Fully Connected Feed Forward ANN

Weight matrices represent connections between neurons in successive layers, and each neuron in the input bias is applied to the ANN, which includes output neurons but excludes input neurons. Where b^l_j signify the bias of neuron j in layer$l$. The weight between neuron k in layer l-1 and neuron j in layer l is denoted byw^l_{jk}. Regardless of the method of activation, the layer l will be represented by the activation functionσ_l. Assume that the activation function is only used for difficulty in each layer. Where j in layer is signifies as output layer l will be denoted byy^l_j. In figure 6 represents the diagram of fully connected feedforward artificial neural network [15].

The weighted input, which is definite as a quantity that will be used a lot,

$$z^l_j = \sum_k w^l_{jk} \sigma_{l-1}(z^{l-1}_k) + b^l_j \qquad (7)$$

All inputs are taken towards the sum to neuron j in layer l. More number of neurons present in the layer which represents l-1. In fact, the weighted input (7) alternatively represented in relations of the previous layer's output as

$$z^l{}_j = \sum_k w^l{}_{jk} y^{l-1}{}_k + b^l{}_j \qquad (8)$$

Where the output $y^{l-1}{}_k = \sigma_{l-1}(z^{l-1}{}_k)$ is the process of weight input activation. Formula (7) explains about working with the deep ANN, It easily explains past weighted inputs through the ANN in terms of recursion and definition are

$$\sigma_0(z^0{}_j) = y^0{}_j = x_j \qquad (9)$$

Which terminates any recursion.

6. **Back propagation:** Given some data $x = [x_1, \dots . x_N]^T \in \mathbb{R}^N$ and some target outputs $y = [y_1, \dots . y_N]^T \in \mathbb{R}^M$ request to select the weight and biases such that $y^L(x; w, b)$ is a good approximation of $y(x)$. The notation $y^L(x; w, b)$ is use to identify the ANN takes x as input and parametrized by the biases and weights w, b. Weight and biases are used to find the cost function $C = C(y, y^L) : \mathbb{R}^N \to \mathbb{R}$ and calculate

$$w^*, b^* = arg_{w,b} \min C\ (y, y^L). \qquad (10)$$

Optimization approaches are different, compared gradient-based and gradient-free, this can be used to explain the minimization problem (10).

V. PREDICTION OF MATERIAL PROPERTIES USING SUPERVISED LEARNING MODELS

In machine learning algorithm, selected as a specific in significant step in the development of machine learning system, since a significant impact on prediction accuracy and generalization ability. There is no algorithm that is ideal for all issues because each algorithm has its own scope of use. As shown in Figure 6, machine learning algorithms has four types that typically employed in materials properties: clustering, regression, classification and exactly, probability estimation, algorithms are mainly used for new regression, classification, materials discovery, and On the micro- and macro-levels, clustering techniques are utilized to prediction of material properties [16].
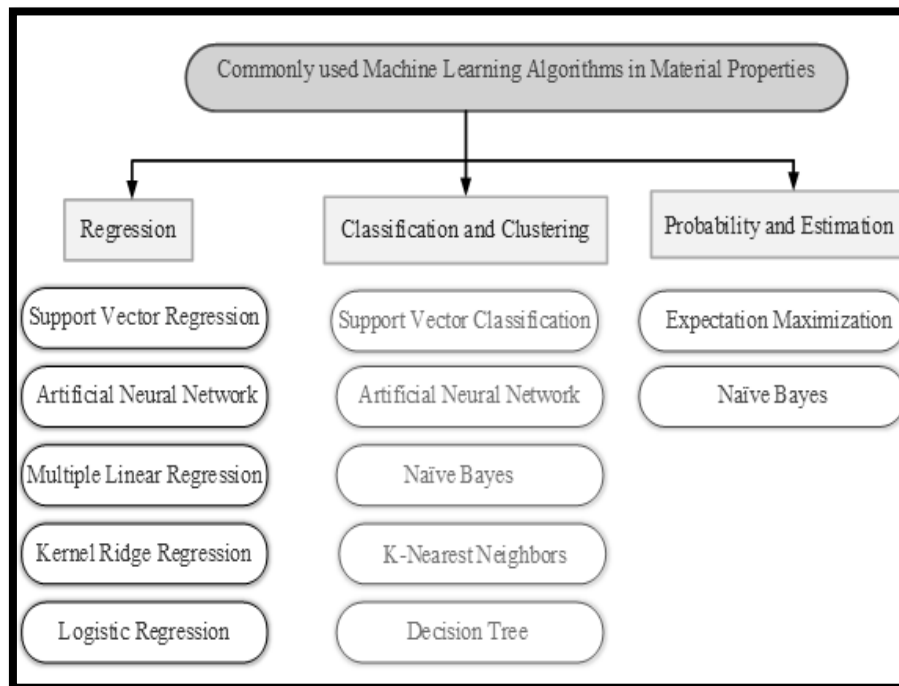
Figure 7: Commonly used Machine Learning Algorithms in Material Properties

VI. THE APPLICATION OF MACHINE LEARNING IN MATERIAL PROPERTY PREDICTION

Material property, including such as glass transition temperature, melting point, hardness, lattice constant, molecular atomization energy, and ionic conductivity, can be described at either the microscopic level or macroscopic. The material properties are learning two corporate properties: experimental measurement and computational simulation. These two approaches contain experimental setup and complicated operations. Consequently, developing computational simulations that adequately represent, numerous logical connections between a material's qualities and its related components is extremely challenging, as well as some of the connections are unknown. Furthermore, experiments to measure compound attributes are normally done at the end of the materials selection phase. Consequently, if the results aren't appropriate, all of the effort and experimental resources invested from that point for nothing. Then, even with large computational or experimental efforts, studying the characteristics of materials is difficult or practically impossible. Then, high-performance prediction models and intelligence are essential for precisely forecasting material properties at a computational cost and low temporal. The advancement of machine learning and the comprehension of algorithms capable of removing patterns from data. The central concept behind utilizing machine learning methods to forecast material properties is to examine and map the correlations (usually nonlinear) between a material's qualities and their relevant components by extracting information from current empirical data.

Machine learning approaches are successful predicting a wide range of material properties. Introduction of various properties that were predicted can be found. In the following, section discuss about some properties in prediction methods, analyzed in different works, which provide good examples for current problems in computational materials properties, and possible strategies to overcome them. The above section explained the ML

application in material property prediction four types of techniques used in material property was briefly explained below.

1. **SVM technique for material property prediction:** The SVM approach is a supervised learning method based on SLT is used to analyses data and recognize patterns for regression and classification. SVM models, are to ANNs, are more suited to limited sample sizes and overcome problems are grateful "the curse of dimensionality" and "overlearning". It shows many exclusive, benefits for resolving high-dimensional problems and nonlinear. By integrating support vector regression (SVR) and genetic algorithm (GA), a hybrid methodology proficient of predicting the atmospheric corrosion behaviour of metallic materials like steel and zinc has been developed. The findings show that this hybrid model outperforms previous techniques in terms of prediction, making viable alternative tool for predicting zinc and steel atmospheric corrosion behaviours. To improve upon a single SVR model's poor performance in predicting the temperature of the As-Se glass transition, involves drastic structural variations, Feature Selection based Two-Stage SVR (FSTS-SVR) using a predicting approach based on structural analysis to investigate the turning point and create each stage prediction model.

2. **Random forest technique for material property prediction:** Random Forest is a supervised learning approach for solving classification problems and regression. It's an ensemble algorithm that creates a class that represents, categorization mode of the individual trees after training a large number of decision trees. When combined with composition features, RF classifiers have demonstrated good prediction performance in earlier studies. Between these models and machine learning algorithms, Random Forest (RF) models have shown promising potential in forecasting a number of material properties, such as superconducting materials' critical temperatures and the capacity of a particular composition to melt spin an amorphous ribbon of metallic glass [17].

   To determine the bulk modulus, an RF regression model was used. RF models are also commonly employed in other fields of research. For example, when compared to various methods RF based methods are good, other approaches are automatically finding the molecular description for nuclear hormone receptor ligands and kinase. Recently, a model of deep learning-based neural network have shown to be extremely successful in applications including as robotics, autonomous driving, image recognition and automatic machine translation, Essentially, the success of material discovery and some difficulties including such as superconductor critical temperatures and prediction make it interesting future materials property applications.

3. **ANN Technique for Material Property Prediction:** ANNs are a sort of model in machine learning that is inspired by biological neural networks is used to estimate or functions with a large number of unknown inputs that are approximate. ANN technique is a nonlinear statistical analysis tool has a high degree of adaptability and self-learning. Back propagation ANNs (BP-ANNs) mainly used to predict the behavior of materials such as compressive properties, corrosion, temperature responses elongation, tensile, and wastage.

   Estimation of polymer glass transition temperatures, researchers used linear regression model and BP-ANN, and indicated that the former's average prediction error (17 K) was significantly lower than the latter's (30 K). By permitting acceptable

prediction errors to be used to solve such problems, BP-ANNs can provide significant advantages business and a fair generalization capacity because no physical examination background information is required. Though, a moderate rate of convergence and may typically decrease into limited minima. Otherwise, RBF-ANNs are a type of ANN effectively overcomes the problem of local minima while also delivering a high convergence rate by integrating the ANN principle of radial basis function [18]. An RBF-ANN is used to investigate fracture propagation in a bituminous layered pavement construction as the thickness of bituminous layer B2, cracking rises considerably.

Furthermore, other applications of ANN modelling include predicting melting points, biofuel compound density and diffusion barriers, excited-state energies, viscosity, and other functional features. The ability to employ ANNs as an arbitrary function approximation method that 'learns' from observed data is its fundamental advantage. An advantage of ANN model has requiring little previous knowledge of the target material and allows for precise learning of the rules regulating property variations from empirical data. But, one of the most common concerns about the problem with ANNs need a large and diversified training dataset with enough sample instances for property prediction, in order to arrest the original structure and apply their discoveries to new scenarios. Moreover, the gained knowledge is hidden in an enormous number of neural network connections, resulting in low comprehensibility, i.e., poor knowledge transparency and explain ability.

4. **Naïve Bayesian Technique for Material Property Prediction:** The Naive Bayes classifier is a quick and eager learning classifier. As a result, real-time predictions might be made with it. Multi class Prediction: This algorithm is well recognized for its ability to predict many classes. Predict the probability of numerous classes of target variables in this situation. Data classification is a supervised machine learning method in the context of machine learning. The training data set for a classification problem is made up of data from multiple experiments. The new data includes both descriptive and target features. The descriptive features are those that have an immediate impact on the target feature's outcome [19].

Each experiment's descriptive characteristics are represented by a vector ($x = x_1, x_2 \ldots x_n$), where $x_1 \ldots n$ represents the values of n descriptive features. Moment, $x_1 \ldots n$ it is one of the following data types, it is advantageous to note that: categorical, numeric values, binary, textual or ordinal. In the event of categorization issues, each feature vector is expressive x has an output y, this is usually a class label, that is $y \epsilon \{C_1, C_2 \ldots C_m\}$ where $C_1, C_2 \ldots C_m$ are $'m'$ class labels or the outcomes of experiments can be classified into simple class. A supervised machine learning-based classification issues main purpose train predictive models utilizing training data from experiments with classification labels, and use to experiments with no classification labels, the learned prediction models are used to forecast classification labels.

In the machine learning domain, among the most commonly used classifiers are probabilistic classification algorithms. The classification algorithm is significant may be attributed to the particular benefits they offer the classification approach is over non-probabilistic. Next things, the algorithm of probabilistic classification deliver the measurements of probability each class that the experiment could decrease into, providing

the end user with more information. The probabilistic classification techniques may be applied to a selection of machine learning responsibilities without creating errors, different material qualities are different.

VII. PERFORMANCE OF DIFFERENT SUPERVISED LEARNING MODEL FOR MATERIAL PROPERTIES PREDICTION

ML approaches have been utilized as a new strategy by various researchers to forecast the material property strength of concrete materials. The experimental design of material property prediction tested and validated on MATLAB. Four various machine learning approaches such as accuracy, precision, specificity and sensitivity used to predict material property. The performance matrix of these four techniques is considered as NB (Naïve Bayesian), SVM (Support Vector Machine), RF (Random Forest) and ANN (Artificial Neural Network),

Table 1: Prediction Performance of Various ML Techniques

Technique	No of Properties	Name of Properties	Accuracy
ANN [16]	3	Hardness, Melting, Strength	98
SVM [20]	2	Strength, Density	97
RF [17]	2	Density, Chemical Stability	96
NB [19]	2	Frequency, Density	97

Prediction performance of various ML techniques is shown in table 1 four different techniques are used to predict the material properties. Three different properties are used in ANN technique that is hardness, melting and strength. Two different properties that are used in SVM technique that is strength, and density. Two different properties that are used in RF technique that is density and chemical stability. Final, two different properties that are used in NB technique that is frequency, and density. The accuracy reached by using ANN is 98%, SVM is 97%, RF is 96% and NB is 97%.

Table 2: Execution time of Various Techniques

Technique	Execution Time
ANN	30 Seconds
SVM	35 Seconds
RF	37 Seconds
NB	39 Seconds

Execution time of Various Techniques is shown in table 2 execution time of material prediction by using ANN is 30 seconds, SVM is 35 seconds, RF is 37 seconds and NB is 39 seconds.
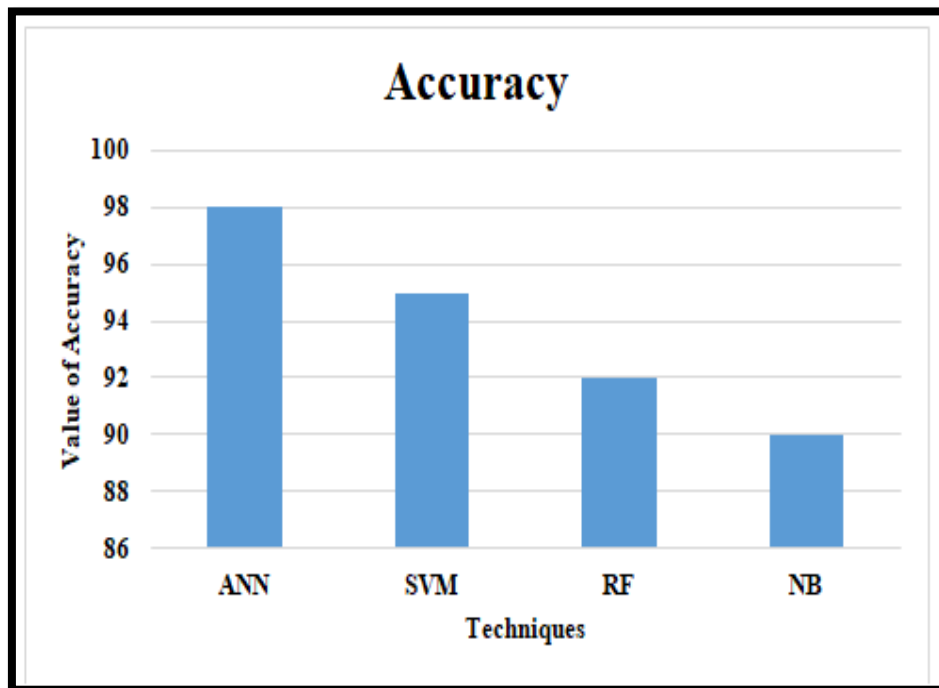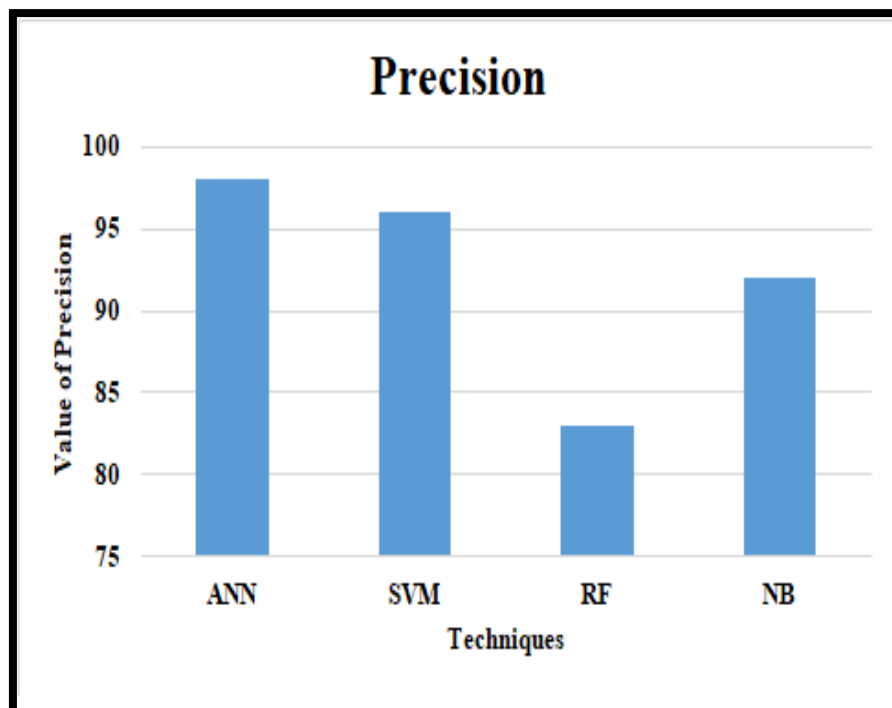
Figure 8: Accuracy of Material Property

Figure 9: Precision of Material Property

Figure 8 represents the accuracy of material property. The graph is plotted in different techniques on X-axis and obtained accuracy value on Y-axis respectively. The comparison of the accuracy value for ANN, SVM, RF and NB techniques are 98, 95, 92 and 90 respectively. This Comparison shows that the ANN technique give better performance compared to other

existing techniques. Figure 9 represents the precision of material property. The graph is plotted in different techniques on X-axis and obtained accuracy value on Y-axis respectively. The comparison of the precision value for ANN, SVM, RF and NB techniques are 97, 96, 83 and 92 respectively. This Comparison shows that the ANN technique give better performance compared to other existing techniques.
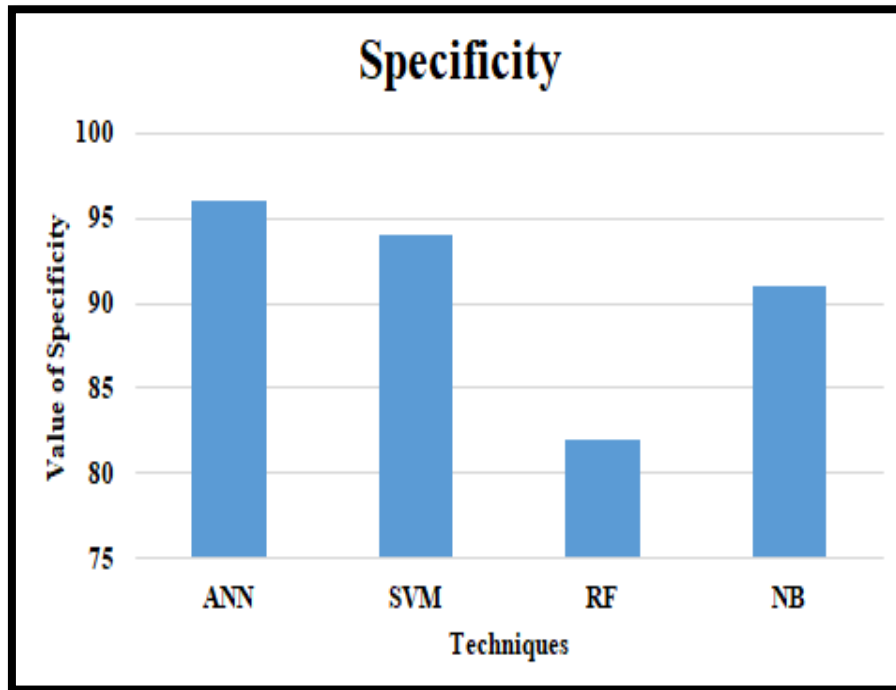

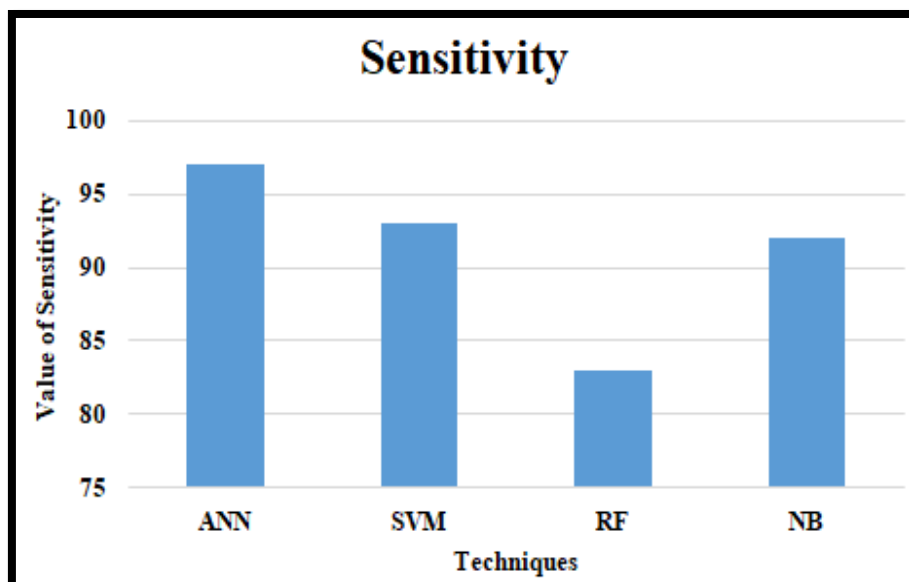**Figure 10: Specificity of Material Property**

Figure 11: Sensitivity of Material Property

Figure 10 represents the specificity of material property. The graph is plotted in different techniques on X-axis and obtained accuracy value on Y-axis respectively. The comparison of the accuracy value for ANN, SVM, RF and NB techniques are 96, 94, 83 and 92 respectively. This Comparison shows that the ANN technique give better performance compared to other existing techniques. Figure 11 represents the sensitivity of material property. The graph is plotted in different techniques on X-axis and obtained accuracy value on Y-axis respectively. The comparison of the sensitivity value for ANN, SVM, RF and NB techniques are 97, 94, 84 and 92 respectively. This Comparison shows that the ANN technique give better performance compared to other existing techniques.

VIII. SUMMARY

In this chapter the discussion is about prediction of material property in Machine learning approach. In the first section, the overview about the material property were explained. After that, brief introduction and some types of material property were discussed. In addition, purpose of predicting material properties also illustrated well. Moreover, brief explanation of supervised learning models and related architecture is conversed. Following that, types of supervised learning models and related diagrams are briefly explained in this section. Furthermore, machine learning algorithm for predict the material property in this section, several supervised learning models were used in the prediction of material properties such as, Naïve Bayesian (NB), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), were discussed. Finally, four techniques are used to predict the material property discussed and its performance also validated in the above section. The ML technique is used to deliver well-organized tools for the properties of the material, as demonstrated by a wide range of relevant studies. The machine learning techniques used in materials properties become larger as theories and methodologies continue to improve.

REFERENCE

[1] Stoll, A., & Benner, P. (2021). Machine learning for material characterization with an application for predicting mechanical properties. *GAMM-Mitteilungen*, *44*(1), e202100003.

[2] Kumar, D., Alam, M., Zou, P. X., Sanjayan, J. G., & Memon, R. A. (2020). Comparative analysis of building insulation material properties and performance. *Renewable and Sustainable Energy Reviews*, *131*, 110038.

[3] Su, M. N., & Young, B. (2019). Material properties of normal and high strength aluminium alloys at elevated temperatures. *Thin-walled structures*, *137*, 463-471.

[4] Deng, B., & Zhang, Y. (2020). Critical feature space for predicting the glass forming ability of metallic alloys revealed by machine learning. *Chemical Physics*, *538*, 110898.

[5] Lyu, X., Shu, G. P., & Du, E. F. (2020). Parameters effect on predicting fire resistance of ultra-high strength concrete filled protected square steel tubular columns. *International Journal of Steel Structures*, *20*(5), 1783-1795.

[6] Zhou, X., Liu, H., Pourpanah, F., Zeng, T., & Wang, X. (2021). A Survey on Epistemic (Model) Uncertainty in Supervised Learning: Recent Advances and Applications. *Neurocomputing*.

[7] Bagherzadeh, J., & Asil, H. (2019). A review of various semi-supervised learning models with a deep learning and memory approach. *Iran Journal of Computer Science*, *2*(2), 65-80

[8]   Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, *4*, 51-62.

[9]   Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, *48*(3), 128-138.

[10] Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE access*, *6*, 33789-33795.

[11] Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel behaviour and society*, *14*, 1-10.

[12] Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, *11*(1), 66.

[13] Tang, J., Chen, X., Hu, Z., Zong, F., Han, C., & Li, L. (2019). Traffic flow prediction based on combination of support vector machine and data denoising schemes. *Physica A: Statistical Mechanics and its Applications*, *534*, 120642.

[14] Arboleda, E. R., Fajardo, A. C., & Medina, R. P. (2018, May). Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors. In *2018 IEEE International Conference on Innovative Research and Development (ICIRD)* (pp. 1-5). IEEE.

[15] Berg, J., & Nyström, K. (2018). A unified deep artificial neural network approach to partial differential equations in complex geometries. *Neurocomputing*, *317*, 28-41.

[16] Liu, Y., Zhao, T., Ju, W., & Shi, S. (2017). Materials discovery and design using machine learning. *Journal of Materiomics*, *3*(3), 159-177.

[17] Song, Y., Lindsay, J., Zhao, Y., Nasiri, A., Louis, S. Y., Ling, J. ... & Hu, J. (2020). Machine Learning based prediction of noncentrosymmetric crystal materials. *Computational Materials Science*, *183*, 109792.

[18] Li, B., Lee, Y., Yao, W., Lu, Y., & Fan, X. (2020). Development and application of ANN model for property prediction of supercritical kerosene. *Computers & Fluids*, *209*, 104665.

[19] Naik, D. L., & Kiran, R. (2018). Naïve Bayes classifier, multivariate linear regression and experimental testing for classification and characterization of wheat straw based on mechanical properties. *Industrial Crops and Products*, *112*, 434-448.

[20] Chaabene, W. B., Flah, M., & Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, *260*, 119889.