International Journal of Research in Engineering, Science and Management
Volume 7, Issue 11, November 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Low Alloy Steel Tensile Properties Prediction Using Machine Learning

Deepraj Chetan Patil[1*], Darshan Devendra Kawade[2]
[1,2]*Student, Dr. D. Y. Patil Institute of Engineering, Management & Research, Akurdi, Pune, India*

***Abstract*: In the current version of industrialization i.e. industry 4.0, implementation of AIML (Artificial Intelligence & Machine Learning) and Automation and Robotics are at the centre of the industries. Here in this project, we are implementing the concepts of ML (Machine Learning) to determine the tensile properties of low alloy steel which is one of the most used materials in the manufacturing industries. Generally, for determining and evaluating the properties of any type of steel traditional method of using UTM (Universal Testing Machine) is preferred, UTM itself is a piece of very costly equipment and it is a destructive type of testing method which are time consuming, labor- intensive and causes wastages of material. So to overcome these cons of UTM methods here we are using data science and machine learning to predict the properties of steel depending upon the composition of its alloying elements. With the help of available standardized data, we are predicting four properties which are Tensile Strength (MPa), 0.2% proof stress (MPa), Elongation (%), and Reduction in Area (%).**

***Keywords*: Low alloy steel, Machine Learning, Tensile Strength (MPa), 0.2% proof stress (MPa), Elongation (%), Reduction in Area (%), Linear Regression, Random Forest, SVM (Support Vector Machine), Decision Tree.**

1. Introduction

This is a machine learning based project to predict the tensile properties of low alloy steel which are generally determined by tensile testing, which is time intensive and requires labour work up to a certain extent. To estimate the tensile properties of steel destructive testing is preferred, which is done with the help of a traditional UTM (Universal Testing Machine). This traditional approach is costly and time-consuming and it is not eligible for small-scale industries to afford UTM.

A. Aim

The primary objective of this project is to leverage data science and machine learning techniques to forecast the tensile properties of low alloy materials by utilizing standardized data on the weight/weight composition of different alloying elements at a specific temperature, along with corresponding estimated tensile properties. This dataset was made available by the NIMS (National Institute of Material Science).

The dataset consists of 19 columns in total among which 15 are input parameters and 4 are output parameters which are the tensile properties to be predicted. It contains 916 rows of data in total. Each alloy has multiple entries with variations in the testing temperature.

Input parameters are,

1) Percentage W/W of
   a. Carbon
   b. Silicon
   c. Manganese
   d. Phosphorus
   e. Sulphur
   f. Nickel
   g. Chromium
   h. Molybdenum
   i. Copper
   j. Vanadium
   k. Aluminium
   l. Nitrogen
   m. Niobium
   n. Tantalum
2) Ceq (Carbon Equivalent Content)
3) Temperature in degrees Celsius (°C).
4) The output parameters to be predicted:
   a. Tensile Strength (MPa)
   b. 0.2% Proof Stress (MPa)
   c. Elongation (%)
   d. Reduction in Area (%).

By employing diverse models and algorithms, a substantial level of accuracy in prediction can be attained to a certain degree. This project incorporates a total of four algorithms and conducts predictions through eight models. The algorithms utilized encompass Linear Regression, Decision Tree, Random Forest and Support Vector Machine (SVM). Principal Component Analysis (PCA) is used as feature extraction method. The eight models employed are as follows: Linear Regression, Decision Tree, Random Forest, SVM (without PCA), as well as Linear Regression, Decision Tree, Random Forest, SVM (with PCA).

The results obtained are satisfactory and reliable with the highest accuracy of 95.54 in predicting Tensile strength via the Decision Tree Model With PCA.

2. Methodology

The data pre-processing step involved cleaning and checking

*Corresponding author: deeprajpatil1010@gmail.com

for skewness, null values or anomalous data. Any data that deviated from a normal distribution was fixed, outliers were either rectified or removed, and the range of independent variables was normalized. Subsequently, the dataset was partitioned into two segments, wherein one portion was allocated for the purpose of training the Machine Learning (ML) model, while the remaining portion served as a testing set to measure the performance of the trained model, maintaining an 80%-20% ratio.

To identify the most suitable ML algorithm for the dataset, four different techniques were tested:

1) Linear Regression

Linear regression is a statistical modelling approach that seeks to determine a linear association between a response variable and one or multiple predictor variables. It assumes a linear equation in the following format:

$$Y = \beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta n Xn$$

where, the line is determined by an intercept (β0) and coefficients (β1, β2, ..., βn) assigned to each independent variable (X1, X2, ..., Xn). The goal of linear regression is to figure out the most accurate values for these coefficients, allowing us to make predictions or draw conclusions about how the variables are related. This is done by minimizing the difference between the observed values and the values predicted by the line.

2) Random Forest

Random Forest is a highly effective algorithm for making predictions by combining multiple decision trees. It finds extensive application across diverse domains such as data science and machine learning.

In a Random Forest, a group of decision trees or a forest is built, with each tree trained on a distinct subset of the data and using a random subset of features. During the prediction phase, each tree in the forest independently makes its own prediction. The final prediction is then determined by either taking the majority vote which is used for classification or averaging the individual tree predictions which is used for regression.

The key idea behind Random Forest is that by combining multiple decision trees, the overall predictive performance can be improved. The algorithm leverages the concept of "wisdom of the crowd," where the collective decision of multiple trees tends to be more accurate and robust than the decision of a single tree.

Random Forest has several advantages such as accommodating various types of features, mitigating overfitting and handling high-dimensional data. It can handle both regression and classification tasks, making it a versatile algorithm.

Furthermore, Random Forest offers the capability to assess the importance of different features in the prediction process. This means it can provide insights into which features have the most influence on the outcome. This information is valuable for understanding the underlying factors driving the predictions.

In summary, Random Forest is widely recognized as a popular and powerful algorithm for predictive modelling. It is known for its ability to deliver precise and dependable predictions in a wide range of domains. Its ability to provide feature importance measures adds to its appeal and usefulness in practical applications.

3) Decision Tree

A decision tree is a flexible machine learning algorithm that is commonly used for both classification and regression tasks. It adopts a treeshaped structure in which each internal node denotes a feature whereas, each branch corresponds to a outcome of that feature, and each leaf node represents a predicted value.

The decision tree algorithm learns from the data by partitioning the feature space recursively as to make smaller and smaller set of data based on the selected features and their associated values. The partitioning is done in a way that optimizes certain criteria, such as maximizing information gain or minimizing impurity, depending on the specific algorithm variant used.

During the prediction phase, a new data instance is traversed through the decision tree from the root to a leaf node, based on the values of its features. The class label or predicted value associated with the reached leaf node is then assigned to the instance.

4) Support Vector Machine (SVM)

Support Vector Machines (SVM) are robust and adaptable machine learning algorithms extensively utilized for both classification and regression purposes. Their strength lies in effectively handling complex datasets with numerous dimensions.

The primary objective of an SVM is to find an optimal “hyperplane" that separates data points of different classes with the maximum margin. In binary classification, this hyperplane acts as a decision boundary, where data points on one side are assigned to one class, and points on the other side belong to the other class. SVMs have several advantages, including their ability to handle high dimensional data, resistance to overfitting, and effectiveness in dealing with small to moderate-sized datasets.

They also have a solid theoretical foundation and offer good interpretability, as support vectors can provide insights into the classification process.

However, SVMs may be computationally expensive for large datasets, and they can be sensitive to the choice of hyperparameters and the kernel function.

Additionally, SVMs are primarily binary classifiers, but they can be extended to handle multi-class problems using techniques such as one-vs-one or one-vs-all.

Each of these algorithms was tested without the use of Principal Component Analysis (PCA) to establish a baseline for comparison. Following this, Principal Component Analysis (PCA) was used to identify the input parameters that were contributing the most to the properties to be predicted for each output parameter. To assess the impact of Principal Component analysis (PCA) on the Machine Learning Models, the analysis was conducted both with and without PCA.

The four previously tested algorithms i.e., Linear Regression, Random Forest, Decision Tree, and Support Vector Machine (SVM) were again evaluated, this time with the inclusion of PCA. The Principal Components found for each of the output

parameters were used in the evaluation of each model. Then the accuracy of each model was assessed, and the results were compared to identify the most suitable algorithm for the given dataset and for each output parameter.

There are four output parameters that are to be predicted by the machine learning models:

1) Tensile Strength

Tensile strength is the maximum amount of stress that a material can endure without fracturing or tearing when subjected to stretching or pulling forces. In brittle materials, such as certain ceramics, the ultimate tensile strength is typically near the point when the material begins to yield or deform. On the other hand, in ductile materials like most metals, the ultimate tensile strength can often surpass at the yield point, allowing for greater resistance to deformation before failure occurs.

Tensile strength is an important mechanical property that provides information about the material's capability to withstand forces acting in opposite directions along its length. It is commonly measured through experiments involving the gradual application of tension until the material breaks. This parameter plays a pivotal role in engineering and structural design, as it determines the load carrying capacity and reliability of components under tensile loads.

2) 0.2% Proof Stress

0.2% proof stress, also known as the offset yield strength or yield point, is a mechanical property used to measure the strength of a material. The 0.2% proof stress specifically pertains to the stress level at which a material starts to display permanent deformation or plasticity. Unlike the ultimate tensile strength, which represents the maximum stress a material can endure before failure, the 0.2% proof stress emphasizes the point at which yielding begins.

It is typically determined by gradually increasing the load applied to a specimen and monitoring its corresponding deformation. The stress at which the material exhibits 0.2% strain is considered 0.2% proof stress. Its stress is commonly used in the design of safety-critical structures and components, such as bridges, buildings, and pressure vessels.

3) Elongation

Elongation is an important mechanical property used to assess the ductility and deformation characteristics of a material. It measures the extent to which a material can stretch or elongate before it breaks or fractures under tensile stress.

Elongation is typically a percentage and represents the delta in the length of a specimen when subjected to a tensile force. It is determined by measuring the original length of the specimen before applying the load and comparing it to the length at the point of fracture or failure.

Elongation is a significant property in materials science that measures a material's ability to deform plastically before fracture. It is a critical factor in assessing ductility, toughness, and overall mechanical behavior. By understanding the elongation characteristics of different materials, we can make logical decisions regarding material selection, process optimization, and design considerations to ensure the appropriate performance and reliability of structures and components.

4) Reduction in Area

Reduction in the area is a mechanical property that measures the extent of localized deformation and necking that occurs in a material during tensile testing. It provides important insights into the ductility and deformation of a material under load. Reduction in the area is determined by comparing the original cross-sectional area of a tensile specimen with the area at the point of fracture or failure. It is expressed as a percentage and calculated as the difference between the original area and the fractured area, divided by the original area, multiplied by 100.

It is closely related to the elongation property discussed above. While elongation measures the increase in length of a material specimen before failure, reduction in the area focuses on the decrease in cross-sectional area. It is particularly useful for assessing the localized deformation and necking that occur in materials under tension.

Reduction in the area is a significant property in material science that quantifies the localized deformation and necking observed in a material during tensile testing. It provides insights into the ductility and deformation behaviour of materials under load. By understanding the reduction in area characteristics of different materials, engineers can make informed decisions regarding material selection, process optimization, and design considerations to ensure appropriate performance and reliability in various applications.

5) Data Collection and Pre-processing

The data used in this project is standardized data made available by the National Institute of Material Science (NIMS) on Kaggle.com. It contains 19 columns along with a column of Alloy Code and 916 rows of data, in which 15 are input parameters and the rest are the output parameters i.e., properties to be predicted.

The data we are using includes the following features:

1. Alloy Code
2. C - Carbon
3. Si - Silicon
4. Mn - Manganese
5. P - Phosphorus
6. S - Sulphur
7. Ni - Nickel
8. Cr - Chromium
9. Mo - Molybdenum
10. Cu - Copper
11. V - Vanadium
12. Al - Aluminium
13. N - Nitrogen
14. Ceq
15. Nb + Ta
16. Temperature (°C)
17. Tensile Strength (MPa)
18. 0.2% Proof Stress (MPa)
19. Elongation (%)
20. Reduction in Area (%)

After importing the dataset, it is checked for any NaN or Null values present in it and it was found that the dataset doesn't contain any null values. After that, the Alloy code column was

Table 1
The accuracy was achieved without PCA implementation

S. No.	Property	Best achieved Accuracy without PCA			
		Random Forest Regression	Linear Regression	Decision Tree	Support Vector Machine
1.	Reduction in the area (%)	82.39	42.67	79.54	59.46
2.	Elongation (%)	86.83	59.83	52.78	54.67
3.	Tensile Strength (MPa)	76	67.43	92.43	68.74
4.	0.2% Proof Stress	90.45	75.54	76.93	66.94

Table 2
The accuracy achieved with the implementation of PCA

S. No.	Property	Best achieved Accuracy with PCA			
		Random Forest Regression	Linear Regression	Decision Tree	Support Vector Machine
1.	Reduction in the area (%)	75.76	43.64	56.75	62.01
2.	Elongation (%)	88.67	68.34	43.67	52.78
3.	Tensile Strength (MPa)	93.56	68.67	95.54	82.98
4.	0.2%Proof Stress	90.66	58.012	66.5	62.09

removed from the dataset as it is irrelevant information.

Each material column value was visualized using a boxplot and was checked for any major skewness present in it. It was noticed that the Cu (Copper) column had a little higher skewness so its skewness was improved by using its square-root values and then the Cu column was replaced by its square-root values column. Similarly, the Al (Aluminium) column was found to have higher skewness and it was replaced by the log (Al) values column.

While visualizing it was found that Tensile strength had an outlier in which its value was greater than 1000 MPa, hence the row containing this outlier value was dropped from the dataset. In the visualization of Elongation (%) the skewness was a little higher, hence the Elongation Column was replaced by the log (Elongation %) values column. After visualization, skewness improvement and outlier removal the dataset was well organized for further analysis.

Then using the organized, cleaned and prepared data correlation matrix was formed which gives the relation between the columns in terms of proportionality and the effect of other column values on each column. For visualizing this correlation between all the features, a heat map was drawn, it is as given in Fig. 1.

B. Web User-Interface

The Web UI developed in Django provides a user-friendly interface for predicting material properties based on different machine learning models. The UI allows users to input the percentage by weight (w/w) of various additives to an alloy and obtain predictions for different material properties. The models are implemented using the pickle system, which enables the retrieval and utilization of pre- trained models or guesses.
Upon accessing the web UI, users are presented with a clean and intuitive interface. The UI consists of input fields. The UI provides clear instructions or hints to guide users in correctly specifying the additive percentages.

Once the user has entered the desired additive percentages, they can submit the input data for analysis. The Django backend then employs the pickle system to retrieve the pre-trained machine learning models. The models encapsulate the knowledge and patterns learned from training data and are used to make predictions based on the provided input.

After the input is processed, the UI displays the predicted material properties to the user. These properties may include one or more of the desired outputs, depending on the specific models or guesses selected. The UI presents the results in a visually appealing and organized manner, making it easy for users to interpret and utilize the predicted material properties for their intended purposes.

The web UI implemented in Django ensures the security and integrity of user data by employing appropriate measures, such as data validation and sanitization techniques. The user-friendly interface should increase the user base of the predictive models to not just the tech-savvy but also laymen.

3. Results

In this section the prediction accuracy of the chosen models obtained after training the models, followed by a general discussion. The accuracy with and without feature extraction is also compared to get an idea of the feature importance, and the effect dimensionality reduction can do to model accuracy. The model comparison can be seen in the table 1 and 2.

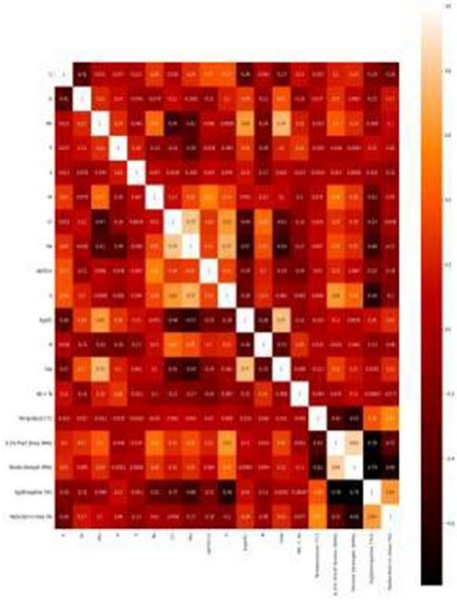

Fig. 1.

4. Conclusion

This paper presented prediction of low alloy steel tensile properties using machine learning.

References

[1] D. Jha, V. Gupta, W. Liao, A. Choudhary, and A. Agrawal, "Moving closer to experimental level materials property prediction using AI," *Scientific Reports*, vol. 12, no. 1, Jul. 2022.

[2] D. Merayo, A. Rodriguez-Prieto, and A. M. Camacho, "Prediction of Physical and Mechanical Properties for Metallic Materials Selection Using Big Data and Artificial Neural Networks," *IEEE Access*, vol. 8, pp. 13444–13456, 2020.

[3] S. Chibani and F.-X. Coudert, "Machine learning approaches for the prediction of materials properties," *APL Materials*, vol. 8, no. 8, p. 080701, Aug. 2020.

[4] E. Bélisle, Z. Huang, S. Le Digabel, and A. E. Gheribi, "Evaluation of machine learning interpolation techniques for prediction of physical properties," *Computational Materials Science*, vol. 98, pp. 170–177, Feb. 2015.

[5] Dr. N. Sandhya*, V. Sowmya, Dr. C. R. Bandaru, and Dr. G. R. Babu, "Prediction of Mechanical Properties of Steel using Data Science Techniques," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 3, pp. 235–241, Sep. 2019.

[6] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, "Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters," *Integrating Materials and Manufacturing Innovation*, vol. 3, no. 1, pp. 90–108, Apr. 2014.

[7] N. S. Reddy, J. Krishnaiah, S.-G. Hong, and J. S. Lee, "Modeling medium carbon steels by using artificial neural networks," *Materials Science and Engineering: A*, vol. 508, no. 1–2, pp. 93–105, May 2009.

[8] S. Chibani and F.-X. Coudert, "Machine learning approaches for the prediction of materials properties," *APL Materials*, vol. 8, no. 8, p. 080701, Aug. 2020.

[9] Y. Wang *et al.*, "Prediction and Analysis of Tensile Properties of Austenitic Stainless Steel Using Artificial Neural Network," *Metals*, vol. 10, no. 2, Feb. 2020.

[10] D. Merayo, A. Rodríguez-Prieto, and A. M. Camacho, "Topological Optimization of Artificial Neural Networks to Estimate Mechanical Properties in Metal Forming Using Machine Learning," *Metals*, vol. 11, no. 8, p. 1289, Aug. 2021.

[11] B. Chanda, P. P. Jana, and J. Das, "A tool to predict the evolution of phase and Young's modulus in high entropy alloys using artificial neural network," *Computational Materials Science*, vol. 197, p. 110619, Sep. 2021.

[12] D. Merayo, A. Rodríguez-Prieto, and A. M. Camacho, "Prediction of Mechanical Properties by Artificial Neural Networks to Characterize the Plastic Behavior of Aluminum Alloys," *Materials*, vol. 13, no. 22, p. 5227, Nov. 2020.

[13] D. Merayo Fernández, A. Rodríguez- Prieto, and A. M. Camacho, "Prediction of the Bilinear Stress-Strain Curve of m Alloys Using Artificial Intelligence and Big Data," *Metals*, vol. 10, no. 7, p. 904, Jul. 2020.

[14] M. Das, G. Das, and M. Ghosh, "Prediction of Mechanical Properties of Sensitized Stainless Steel by Neural Network Modeling and Validation Using Ball Indentation Test," *Journal of Materials Engineering and Performance*, Nov. 2022.

[15] E. Javaheri *et al.*, "Quantifying Mechanical Properties of Automotive Steels with Deep Learning Based Computer Vision Algorithms," *Metals*, vol. 10, no. 2, p. 163, Jan. 2020.

[16] D. M. Dimiduk, E. A. Holm, and S. R. Niezgoda, "Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering," *Integrating Materials and Manufacturing Innovation*, vol. 7, no. 3, pp. 157– 172, Aug. 2018.

[17] G. Partheepan, D. K. Sehgal, and R. K. Pandey, "Quasi-Non-Destructive

[18] Evaluation of Yield Strength Using Neural Networks," *Advances in Artificial Neural Systems*, vol. 2011, pp. 1–8, Jun. 2011.

[19] S. Malinov, W. Sha, and J. J. McKeown, "Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network," *Computational Materials Science*, vol. 21, no. 3, pp. 375–394, Jul. 2001.

[20] A. Eres-Castellanos, I. Toda- Caraballo, A. Latz, F. G. Caballero, and C. Garcia-Mateo, "An integrated- model for austenite yield strength considering the influence of temperature and strain rate in lean steels," *Materials & Design*, vol. 188, p. 108435, Mar. 2020.

[21] O. Sanni, O. Adeleke, K. Ukoba, J. Ren, and T.-C. Jen, "Application of machine learning models to investigate the performance of stainless steel type 904 with *Materials Research and Technology,* vol. 20, pp. 4487–4499, Sep. 2022.

[22] G. Dhande and Z. Shaikh, "Analysis of Epochs in Environment based Neural Networks Speech Recognition System," *in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI),* Apr. 2019.

[23] X. Wu and J. Liu, "A New Early Stopping Algorithm for Improving Neural Network Generalization," in *2009 Second International Conference on Intelligent Computation Technology and Automation*, 2009. Accessed: May 24, 2023.

[24] O. Adeleke, S. A. Akinlabi, T.-C. Jen, and I. Dunmade, "Application of artificial neural networks for predicting the physical composition of municipal solid waste: An assessment of the impact of seasonal variation," *Waste Management & Research: The Journal for a Sustainable Circular Economy*, vol. 39, no. 8, pp. 1058–1068, Feb. 2021.

[25] P. Goyal, *Artificial Intelligence and Machine Learning For SPPU*. Pune: Tech Knowledge, 2022.

[26] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*. Springer, 2019.

[27] T. P. Trappenberg, "Machine learning with sklearn," in *Fundamentals of Machine Learning*, Oxford University Press, 2019, pp. 38–65.

[28] "Simple Linear Regression," in *Applied Linear Regression*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005, pp. 19–46.

[29] NISM, "Mechanical Properties of Low Alloy Steel," *Kaggle*. https://www.kaggle.com/datasets/nitinsharma21/mechanical-properties-of-low-alloysteel

[30] X. Shi, "Nonlinear PCA & Feature Extraction," in *Blind Signal Processing*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 84–96.

[31] J. VanderPlas, *Python Data Science Handbook: Essential Tools for Working with Data*. "O'Reilly Media, Inc.," 2016.