Machine Learning Prediction of Mechanical Properties in Reinforcement Bars: A Data-Driven Approach

Taimoor Hussain Qureshi[1,2], Muhammad Ali Siddiqui[1*], Syed Dabeer Mehdi Naqvi[1,3], Muhammad Zeeshan[1,4], Minhal Waseem[1] and Zain Noreen[1,5]

[1] Computational and Experimental Materials Innovation Group (CEMIG), Department of Metallurgical Engineering, NED University of Engineering and Technology, Karachi, Pakistan

[2] Celeritas Digital, Karachi, Pakistan

[3] IRD, Pakistan

[4] Department of Mathematics, NED University of Engineering and Technology, Karachi, Pakistan

[5] FAST National University of Computer and Emerging Sciences, Karachi Pakistan

* Correspondence: Dr. Muhammad Ali Siddiqui, m.siddiuqi@neduet.edu.pk

Citation | Quraishi. H. T, Siddiqui. M. A, Naqvi. D. M, Zeeshan. M, Waseem. M, Noreen. Z,"Machine Learning Prediction of Mechanical Properties in Reinforcement Bars: A Data-Driven Approach", IJIST, Vol. 6 Issue.2 pp 413-425, April 2024

Received | Mar 27, 2024 **Revised |** Apr 14, 2024 **Accepted |** April 22, 2024 **Published |** April 25, 2024.

Introduction/Importance of Study: This study addresses the pressing need for precise prediction of mechanical properties in steel reinforcement bars (rebars) through a data-driven approach utilizing machine learning techniques.

Novelty statement: Our research provides a solution to the challenge of predicting mechanical properties in rebars using advanced machine learning algorithms, filling a critical gap in existing methodologies.

Material and Method: Our study utilized a meticulously curated dataset comprising over 10,300 samples of diverse rebar types manufactured through industrial methods. We leveraged the latest PyCaret model to integrate machine learning algorithms, with a focus on training and rigorously testing linear regression models. Data preprocessing involved thorough cleaning using Python libraries such as Pandas and NumPy, supplemented by cross-validation techniques to ensure robust model generalization.

Result and Discussion: The core findings of our study revolve around the linear regression model algorithm trained within the machine learning framework, enabling precise determination of key mechanical properties including Yield Strength (YS) and ultimate Tensile Strength (UTS). Additionally, we explored the Ratio of UTS to YS (UTS/YS) as a critical mechanical property, incorporating essential input features such as weight percent of carbon (C), manganese (Mn), silicon (Si), carbon equivalent (Ceq), quenching parameters (Q), and diameter (d).

Concluding Remarks: Our research offers valuable insights into the application of machine learning for the precise prediction of mechanical properties in reinforcement bars, contributing to enhanced quality control and optimization in the steel manufacturing industry.

Keywords: Machine Learning, Reinforcement Bars (Rebars), Data-Driven Approach, Chemical Composition, and Quality Control

Introduction:

Reinforcement bar (rebar), as defined by the American Society for Testing and Materials (ASTM A615M) [1] and British Standard (BS 4449)[2], is an essential material in reinforced concrete structures. Rebar is a crucial component in reinforced concrete structures which primarily comprises iron and carbon, incorporating minor alloying elements such as Mn and Si along with impurities elements S and P within specified limits, exhibiting excellent weldability and ductility. Its ribbed design enhances adhesion with concrete, ensuring robust structural integrity. Availability of various sizes and grades, construction steel bars, wires, and welded fabric serves diverse construction needs, from building foundations to bridges. Rigorous quality assurance measures ensure consistency and reliability, making construction steels like Grade 40, Grade 60, and Grade 75 a cornerstone of modern infrastructure projects. [1].

In the realm of engineering and construction, the mechanical properties of rebar hold a pivotal role in ensuring the durability, safety, and overall performance of concrete structures. Rebar provides the essential tensile strength to concrete, compensating for its inherent weakness in tension. Understanding the mechanical properties of rebar, including its tensile strength, yield strength, and ductility, it is therefore critical in the design and safety assurance of concrete structures. Traditionally, determining the mechanical properties of rebar has relied on experimental testing procedures, such as tensile tests (as per ASTM E8), which can be costly, time-consuming, and potentially lead to wastage of resource. Therefore, developing a reliable method to predict these properties based on readily available information, such as the chemical composition of the rebar, has been a subject of interest for researchers and industry professionals alike [3][4]. Reliable predictions aid in determining the load-bearing capacity, deformation characteristics, and resilience of concrete structures under varying conditions [5].

In recent years, Machine Learning (ML), a branch of Artificial Intelligence (AI), has demonstrated remarkable success in predictive modeling across various engineering domains [6]. ML algorithms learn from data to make predictions or decisions without being explicitly programmed to perform the task. They are capable of handling complex, multi-dimensional data and can often uncover non-linear relationships that might be missed by traditional statistical methods [3]. The advent of ML provides a promising avenue for predicting rebar mechanical properties. Precisely, the application of linear regression models, a basic yet powerful ML technique, offers a method for predicting continuous output variables including the mechanical properties of rebar - based on one or more input variables [3][7].

In this research, we aim to predict the mechanical properties of rebar by utilizing the chemical composition including elements Carbon (C), Manganese (Mn), Silicon (Si), Carbon equivalent (C.E) along with quenching parameters utilized in the Thermomechanical Treatment (TMT) section, such as Water Flow Rate and Rolling Mill Speed. This approach eliminates the need for extensive mechanical testing and offers potential benefits in cost reduction, time efficiency, and enhancement of structural design processes.

Objectives

This study involves collection of real time data on the chemical composition of reinforcement bars (rebars), TMT section, and mechanical testing from an indigenous industry in Karachi, Pakistan. The objective is to develop a machine learning model capable of accurately predicting the mechanical properties of rebars based on their chemical composition and quenching parameters. Additionally, we aim to assess the predictive accuracy and reliability of the developed regression model. This research presents an innovative solution to the problem of predicting mechanical properties in rebars by employing state-of-the-art machine learning algorithms, thus bridging a significant gap in current methodologies.

Proposed Theories:

Regression:

Regression is a statistical method employed primarily for forecasting and prediction, which aligns closely with the applications of machine learning. This involves analyzing data to make predictions about future outcomes based on historical patterns, it also investigate potential causal relationships between independent and specific set of dependent variables within a given dataset[7], [8], [9], [10], [11].

Theory of Linear regression:

Linear regression serves as a foundational statistical approach utilized to depict the connection between a dependent variable, typically denoted as Y, and one or multiple independent variables, often represented as X [8], [9], [10], [11] In the context of this research paper focused on forecasting Yield Strength (YS), Ultimate Tensile Strength (UTS), and ration of Ultimate Tensile Strength to Yield Strength (UTS/YS) using machine learning technique, linear regression is employed to establish the association between the chemical composition of materials (considered as independent variables) and their corresponding mechanical properties (recognized as dependent variables).

Linear Regression Model:

Linear regression entails representing the connection between the dependent variables (such as yield strength, tensile strength, and percentage elongation) and the independent variables (which encompass the chemical composition of elements) through a linear equation.

For Yield Strength (YS):

$$YS = a_0 + a_1C + a_2Mn + a_3Si + a_4C_{eq} + a_5Q + a_6d + e_1 \quad \ldots (1)$$

For Ultimate Tensile strength (UTS):

$$UTS = b_0 + bC + b_2Mn + b_3Si + b_4C_{eq} + b_5Q + b_6d + e_2 \quad \ldots (2)$$

For the Ratio of UTS to YS

$$Ratio = \frac{UTS}{YS} \quad \ldots (3)$$

Where:

$$Carbon\ equivalent(C_{eq}) = C + \frac{Mn}{6} + \frac{Cr+Mo+V}{5} + \frac{Ni+Cu}{15}\ Ref{:}\,[12] \quad \ldots (4)$$

'YS', 'UTS', and the ratio of UTS to YS are the predicted values of the dependent variables. Carbon(C), Manganese (Mn), Silicon (Si), carbon equivalent (Ceq), water flow rate (Q), and diameter (d) are the values of the independent variables.

$a_0, b_0,$ are the intercept terms,

$a_1, a_2, a_3, a_4, a_5, a_6$ or $b_1, b_2, b_3, b_4, b_5, b_6$ are the coefficients (slope) corresponding to each independent variable Carbon(C), Manganese (Mn), Silicon (Si), and carbon equivalent (C.E)respectively.

The e1, or e2, are the error terms representing the difference between the observed and predicted values.

Machine Learning Approach:

Machine learning is an artificial intelligence technique that focuses on creating algorithms capable of finding hidden patterns in data, leading to better predictions or decisions. It involves concepts like supervised learning, unsupervised learning, and reinforcement learning [7], [12]. This approach is advantageous in dealing with multiple independent variables and complex relationships with the dependent variable; it allows the use of algorithms like linear regression to build models without explicitly defining the equation beforehand. In this study, we used supervised learning, where the model learns from labeled data, mapping input to output in order to make predictions on new data. In this study, we utilized the linear regression approach [7]. Many researchers adopt a generalized step-by-step process for predictive modeling. It begins with data collection, where a dataset containing independent variables (features) and the dependent variable (target) is gathered. Following this, data preprocessing is undertaken to clean the data, handle missing values, and potentially scale or normalize features. Subsequently, a linear regression model is trained using the prepared dataset. During this phase, the algorithm adjusts coefficients to minimize the error between predicted and actual values, thereby enhancing its predictive accuracy. To evaluate the model's effectiveness, metrics such as mean squared error or R-squared are utilized on a separate validation dataset or through cross-validation techniques. Once the model's performance meets satisfactory standards, it is deployed for practical use, enabling predictions on new data and supporting decision-making processes. [13].

Experimental approaches:

Machine Learning Approaches:

This research employs machine learning methodologies to predict Yield Strength (YS), Ultimate Tensile Strength (UTS), and the ratio of Ultimate Tensile Strength to Yield Strength (UTS/YS). A linear regression model was utilized to establish the relationship between the chemical composition of materials (considered independent variables) and their mechanical properties (recognized as dependent variables). The model was trained to minimize the sum of squared errors, which quantifies the vertical deviation between observed and predicted values. The machine learning algorithm was meticulously crafted using Python programming language. In this research work, we used PyCaret which is an open-source, low-code machine learning library in Python that automates the end-to-end machine learning process. It provides a simplified interface for various machine learning tasks such as data preprocessing, feature engineering, model selection, hyperparameter tuning, and model evaluation. PyCaret allows users to perform complex machine-learning tasks with just a few lines of code, making it accessible to both beginners and experienced data scientists. Figure 1 represents the flow diagram of experimentation.

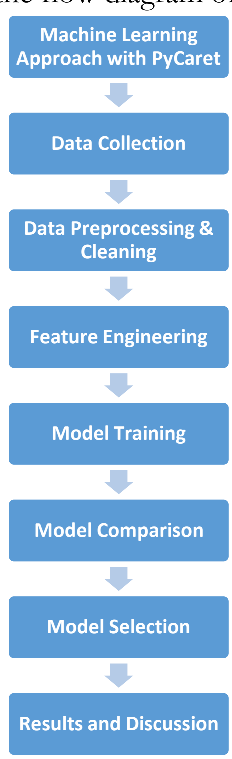

Figure 1. Flow Diagram - Machine Learning Approach with PyCaret for Experimentation

Dataset Collection:

Our dataset consisted of over 10,300 records for this study, selected from the *Indigenous Karachi Steel Industry*, a sector that significantly contributes to the global supply of reinforcement bars for the construction industry. We specifically selected rebars manufactured according to American (ASTM A615) and British (BS4449 standards), ensuring that our sample accurately represents the construction steel industry, including prominent manufacturers such as Faizan Steel, Amreli Steel, Agha Steel, Magna Steel, among others. This deliberate selection aimed to capture a significant market share, thereby enhancing the generalizability of our findings. A summarized version of the dataset is presented in Table. Furthermore, our study utilized standard bar sizes ranging from 16 mm to 36 mm, ensuring a comprehensive analysis of the steel industry's key dimensions.

After meticulously gathering records, as detailed in Table 1, which encompassed input variables such as weight percentage of Carbon (C), Manganese (Mn), Silicon (Si), Carbon equivalent (C.E), as well as critical rolling mill parameters including Water Flow Rate and Rolling Mill Speed. Moreover, our dataset provided comprehensive insights into the factors influencing yield strength, tensile strength, and ratio in steel properties. These dataset values were obtained through real-time analyses conducted via Optical Emission Spectroscopy for chemical composition and tensile tests (as per ASTM standard E8) for mechanical properties on the rebars. Carbon content is known for influencing hardness and strength, while Silicon (Si) content affects heat resistance and oxidation. Manganese (Mn) content enhances both strength and workability. Furthermore, the Carbon Equivalent (C.E) was crucial, reflecting the combined effect of different alloying elements. Additionally, the Thermomechanical Treatment (TMT) parameters, including Mill Speed, impacting grain structure and mechanical behavior, and Water Flow Ratio, affect cooling rates during processing. These parameters were carefully examined for their role in shaping material properties. Lastly, the focus on three critical properties of rebars: Yield Strength (YS), representing the stress at which material exhibits permanent deformation; Ultimate Tensile Strength (UTS), indicating the maximum stress a material can withstand; and the Ratio of UTS/YS, highlighting ductility and toughness, further enriched our understanding of the mechanical characteristics under study.

Table 1. Selected Data Points from the Dataset records used in this Study.

Rebar Size	Sample Weight	C	Si	Mn	C.E	Mill Speed	Water Flow Rate	Yield Strength	Tensile Strength	Ratio
20.0	2.440	0.242	0.230	0.744	0.427	9.70	340	553	675	1.22
20.0	2.452	0.213	0.230	0.749	0.398	9.70	350	553	647	1.17
32.3	6.317	0.222	0.249	0.703	0.383	3.60	310	608	698	1.15
32.3	6.297	0.235	0.194	0.690	0.394	3.60	310	610	720	1.18
32.3	6.292	0.275	0.176	0.654	0.451	3.60	305	589	726	1.23
32.3	6.487	0.275	0.176	0.654	0.451	3.60	280	591	731	1.24
32.3	6.354	0.230	0.199	0.701	0.417	3.60	280	581	695	1.20
32.3	6.418	0.235	0.274	0.780	0.406	3.60	280	595	708	1.19
32.3	6.466	0.226	0.221	0.805	0.418	3.60	280	593	692	1.17
35.8	7.877	0.260	0.267	0.901	0.487	2.80	230	540	657	1.22
35.8	7.903	0.242	0.300	0.964	0.465	2.80	240	582	727	1.25

Data Preprocessing and Cleaning:

Ensuring the accuracy and relevance of our dataset was a critical prerequisite for developing a robust machine-learning model. This process began with data cleaning, which involved handling missing values, managing outliers, and normalizing the data. We utilized Python libraries such as Pandas and NumPy within the PyCaret framework for data preprocessing and cleaning, identifying and appropriately addressing any missing data points through imputation methods or data instance removal based on their nature and quantity. Additionally, PyCaret streamlined these preprocessing tasks, enhancing efficiency and effectiveness in preparing the dataset for model training. We employed the Scikit-learn library's scaling functions for normalization, ensuring that all input feature ranges were within a comparable range to prevent any feature from dominating the model due to its numerical scale.

Model Training with PyCaret:

Our experimental methodology adhered to a rigorous and structured approach, beginning with meticulous curation and purification of the dataset to uphold data integrity. Subsequently, the dataset underwent stratification into distinct training and testing subsets. We evaluated a diverse array of machine learning models, including decision trees, random forests, gradient boosting, support vector machines, and neural networks, utilizing established metrics including Root Mean Squared Error (RMSE), and R-squared [14][13]. The PyCaret library played a crucial role in facilitating model training, validation, and hyperparameter optimization processes. Notably, the linear regression model emerged as the most proficient in navigating and comprehensively addressing the intricate nuances inherent in the dataset's structure and complexities. We conducted a rigorous fine-tuning process for the selected linear regression model, optimizing its parameters, refining its features, and conducting thorough feature selection to enhance its predictive accuracy.

Suitable ML Model Selection with PyCaret:

Moreover, our research also encompassed a comprehensive comparative analysis of diverse machine-learning models aimed to predict the mechanical properties of reinforcement bars. By synthesizing insights from leading researchers, including Murta et al.[.[14], Estela Ruiz et al[13]., Carneiro et al.[15], Stoll & Benner[12], and Mukhopadhyay et al[16], we meticulously evaluated the predictive accuracy and robustness of these models, highlighting the importance of thorough data analysis and meticulous model selection processes in achieving precise and reliable predictions within the domains of materials science and engineering.

Results

We utilized the ML tool PyCaret to process the given dataset via Linear Regression. This outperformed other models such as Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor, and Decision Tree Regressor in terms of Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) values. These metrics indicated that Linear Regression provides a more accurate and reliable prediction of the target variable compared to the alternatives. The findings derived from our linear regression investigations are presented in this section.

Figures 1 and 2 represent scatter plots where the x-axis represents the actual strength (YS and UTS) values of our samples, and the y-axis represents the corresponding predicted strength (YS and UTS) values generated by our machine learning model. Figure 3 illustrates the relationship between the actual ratio of ultimate tensile strength to yield strength (x-axis) and the predicted ratio (y-axis). Ideally, in a perfect prediction scenario, all the data points would fall along a diagonal line (y=x), indicating perfect agreement between the actual and predicted values. The following observations were derived from our analysis.

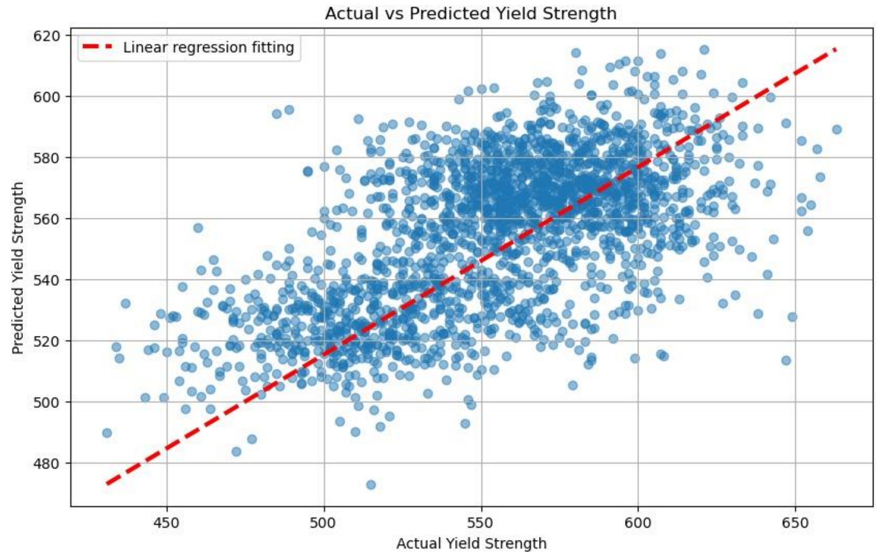

Figure 1: Actual vs. Predicted Yield Strength.

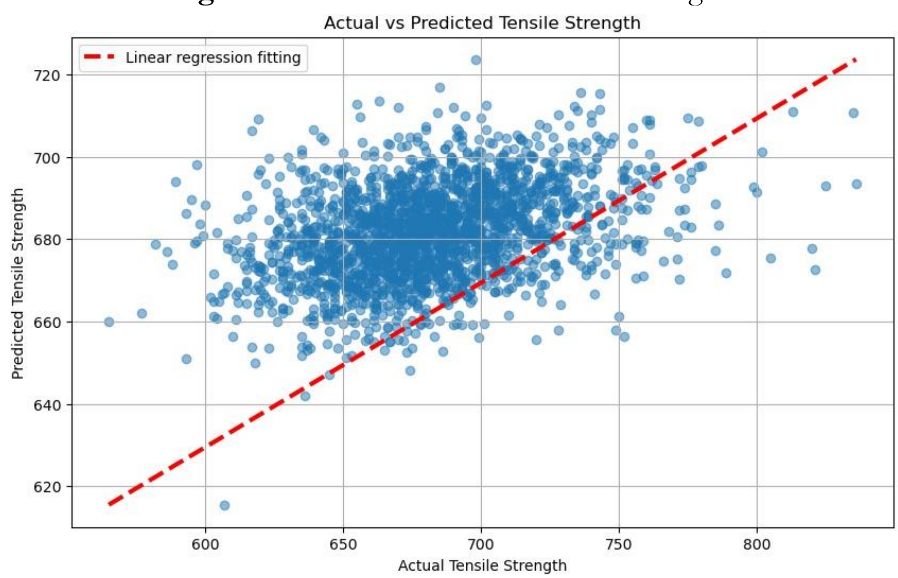

Figure 2: Actual vs. Predicted Tensile Strength.

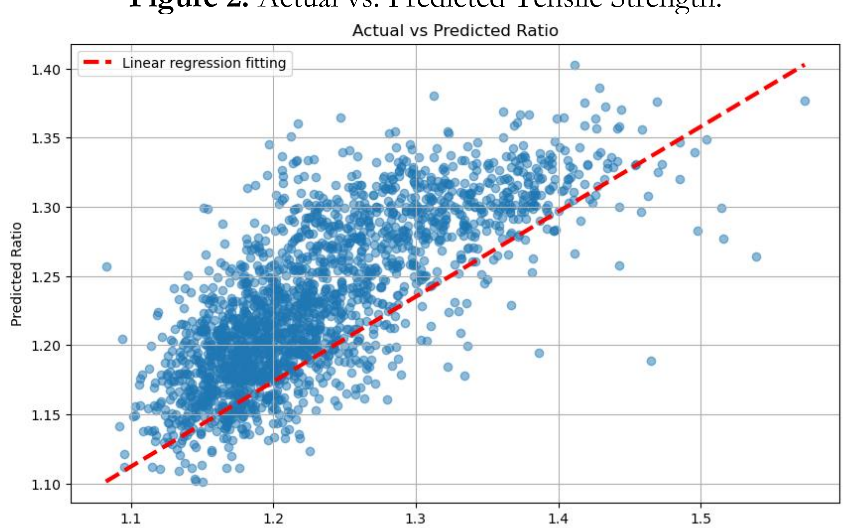

Figure 3: Actual vs. Predicted Tensile to Yield Strength Ratio.

Figure 1 shows Yield Strength, the scatter plot illustrates the relationship between the actual yield strength values (x-axis) and the predicted yield strength values (y-axis) generated by our machine learning model. A tight clustering of points around the diagonal line indicated that the model's predictions closely match the actual yield strength values. The Root Mean Squared Error (RMSE) for Yield Strength was 31.51, indicating the average deviation between the actual and predicted values. The R-squared value of 0.36 suggested that approximately 36% of the variability in the yield strength values was explained by the model.

Furthermore, a cloud of data points was scattered around a diagonal line in the plot, albeit with some deviations from this line. This suggested a positive linear relationship between the actual yield strength and the predicted yield strength. Notably, more data points were concentrated in the middle of the plot in comparison to the edges, indicating that the model tends to make more accurate predictions for yield strengths in the middle range of the data than for yield strengths at the high or low end.

The scatter plot in figure 2 reveals a dispersion of data points around the diagonal line, indicating a positive linear relationship between the actual and predicted tensile strength. This suggests that while the model can predict the general trend in tensile strength, there is some degree of error present. Additionally, the data points appeared to be more densely concentrated in the central region of the plot compared to the periphery. This observation suggested that the model's predictive accuracy may be higher for mid-range tensile strength values compared to values at the lower or higher ends of the spectrum. The RMSE for Tensile Strength was calculated to be 32.65, indicating the average deviation between the actual and predicted values. Furthermore, the R-squared value of 0.13 indicated that approximately 13% of the variability in the tensile strength values was explained by the model.

In figure 3, the scatter plot revealed a positive linear relationship between actual and predicted ratio values. Data points dispersed around the ideal diagonal line, indicating the overall trend but with approximately a 5% margin of error. Notably, a higher concentration of points in the center suggests better predictive accuracy for mid-range ratios. Additionally, a slight downward shift implied a potential overestimation bias, where predictions tend to be higher than actual measurements. The RMSE for the Ratio was 0.05, indicating a small average deviation between the actual and predicted ratios. Moreover, the relatively higher R-squared value of 0.58 suggested that approximately 58% of the variability in the ratio values was explained by the model.

The scatter plots for all three parameters, demonstrate the model's ability to predict the mechanical properties of the steel reinforcement bars accurately. The tight clustering of points around the diagonal line indicates strong predictive performance, while the RMSE and R-squared values provide quantitative measures of the model's accuracy and explanatory power. These results suggested that the machine learning model effectively captures the underlying relationships between the input variables and the mechanical properties of interest.

Discussion

The indigenous steel industry produces a variety of steel products, with Grade 60 (carbon steel) being a prevalent choice for residential, commercial, and large-scale infrastructure projects in Pakistan, following ASTM A615/A615M standards. These industries manufacture steel bars ranging from 10 mm to 40 mm in diameter, with the capability to produce standard-length rebars of 12 meters. This product must meet minimum tensile requirements[1], including a yield strength of 60,000 psi (420 MPa), a tensile strength of 80,000 psi (550 MPa), and a percent elongation of 9%.

The mechanical properties of steel depend strongly on chemical composition and the parameters used during thermomechanical processing [14], which are typically, assessed using a tensile testing machine. However, the challenge of low sampling rates in sample testing presents a significant obstacle to obtaining precise results in laboratory analysis. To address this challenge, statistical tools are being utilized to develop mathematical models via machine learning approaches [14][17], simplifying the prediction process. The forecasted mechanical properties are significantly influenced by several key factors such as the material's chemical composition and mechanized intricacies. Understanding the impact of each variable during production is crucial for achieving high-quality steel production while minimizing costs. Therefore, a mathematical model via ML approach has been created by various researchers, to predict material properties, encompassing yield strength, tensile strength, and percentage elongation.

The research conducted by Murta et al [14] investigated the relationship between chemical composition, heat treatment parameters, and four main mechanical properties of steel rebars: yield strength (YS), ultimate tensile strength (UTS), UTS/YS ratio, and percent elongation. Researchers used linear regression analysis and Artificial Neural Network (ANN) to predict these properties from 18 input variables. The ANN outperformed linear regression in predicting all four properties, as it can capture both linear and nonlinear relationships. Linear regression yielded lower coefficients of determination (R^2) compared to ANN, indicating poorer predictive accuracy.

Tests for significance revealed that carbon, vanadium, and manganese exhibit strong linear relationships with all mechanical properties, consistent with metallurgical literature. The analysis demonstrated a high degree of reliability in the mathematical models, suggesting their potential integration into industrial systems for achieving desired mechanical properties within specification limits and enhancing understanding the influence of constitutional variables and various processing parameters on mechanical properties. Similarly, Multivariate Adaptive Regression Splines (MARS) were employed for predicting the mechanical properties (yield strength, tensile strength, and elongation) of a steel strip. Mukhopadhyay et al [16] introduced a model based on Multivariate Adaptive Regression Splines for predicting the mechanical properties of steel strips. This approach addressed the challenge of accurately predicting properties, which depend on both the chemical composition of the steel and various processing parameters. The model's predictions were found to be in good agreement with actual measured data, indicating its effectiveness in predicting mechanical properties.

Suitable ML Model Selection with PyCaret:

Our research embarked on a comprehensive comparative analysis of diverse machine-learning models aimed at predicting the mechanical properties of reinforcement bars. By synthesizing insights from leading researchers such as Murta et al. [14], Estela Ruiz et al. [13], Carneiro et al. [15], Stoll & Benner [12], and Mukhopadhyay et al. [16], we meticulously evaluated the predictive accuracy and robustness of these models. Through our analysis, we unearthed the superior performance of specific machine learning models, notably highlighting the model proposed by Estela Ruiz et al. [13], in accurately predicting mechanical properties, particularly YS and TS, compared to other models examined [15][12]. This superiority was attributed to the sophisticated algorithms employed, which contributed significantly to enhanced predictive accuracy and reliability. Table 2 shows the comparison of different Machine learning models that were used in the PyCaret ML algorithm. The decision to choose Linear Regression over other regression models is supported by several key metrics and considerations derived from the performance table (Table 2). Linear

Regression outperformed other models such as Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor, and Decision Tree Regressor in terms of Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) values.

Table 2. Comparison of different Machine learning models

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE	TT
Linear Regression	13.0758	421.8065	20.0548	0.9456	0.0933	0.0575	0.31
Random Forest Regressor	13.9428	461.4641	21.195	0.9405	0.0952	0.0601	0.253
Gradient Boosting Regressor	14.8335	477.1152	21.6424	0.9377	0.1061	0.0649	0.116
Extra Trees Regressor	17.5937	574.0868	23.5501	0.9258	0.1031	0.0736	0.742
Decision Tree Regressor	17.4081	733.8058	26.7518	0.9054	0.1164	0.073	0.034

These metrics indicate that Linear Regression provides a more accurate and reliable prediction of the target variable compared to the alternatives. Additionally, Linear Regression demonstrates a lower Root Mean Squared Logarithmic Error (RMSLE) and Mean Absolute Percentage Error (MAPE), further solidifying its superiority in terms of predictive accuracy and consistency.

The Time Taken (TT) metric also favors Linear Regression, showing that it achieves these results efficiently. Overall, based on the comprehensive evaluation of performance metrics, Linear Regression emerges as the optimal choice for this predictive modeling task, offering a balance of accuracy, interpretability, and computational efficiency. The RMSE value of 20.0548 signified the average error between predicted and actual values, with lower values indicating better model accuracy. The R^2 value of 0.9456 indicated that the model explains about 94.56% of the variance in the data, reflecting strong predictive power and a good fit of the model to the dataset. These metrics provided crucial insights into the models' ability to accurately estimate vital mechanical properties such as Yield Strength (YS), Tensile Strength (TS), and the TS/YS ratio, aligning with internationally recognized BS 4449 and ASTM 615 standards.

Influence of Input Variables:

Understanding these nuanced material-process relationships is essential for optimizing production processes and ensuring high-quality steel products. We strategically selected independent variables such as C (Carbon), Mn (Manganese), Si (Silicon), C.E. (Carbon Equivalent), mill speed, and flow rate based on their correlations with dependent variables like yield strength, tensile strength, and steel property ratios. Heatmap analysis (as shown in Figure 4) revealed meaningful relationships: carbon showed moderate positive correlations with yield and tensile strength, consistent with its recognized role in steel strengthening; manganese and silicon exhibited weaker but still positive correlations with these mechanical properties. Additionally, Carbon Equivalent (C.E.) demonstrated a notable positive correlation, emphasizing its relevance in influencing steel hardenability and subsequent strength properties.

While mill speed and flow rate displayed weaker correlations, their inclusion acknowledges their potential influence, albeit with a more nuanced impact requiring further exploration. This analysis substantiates the rationale behind selecting these independent variables, particularly highlighting the importance of chemical composition (C, Mn, Si, C.E.) in determining the mechanical properties of the steel. Thus, it provides a compelling justification for their inclusion in the study.

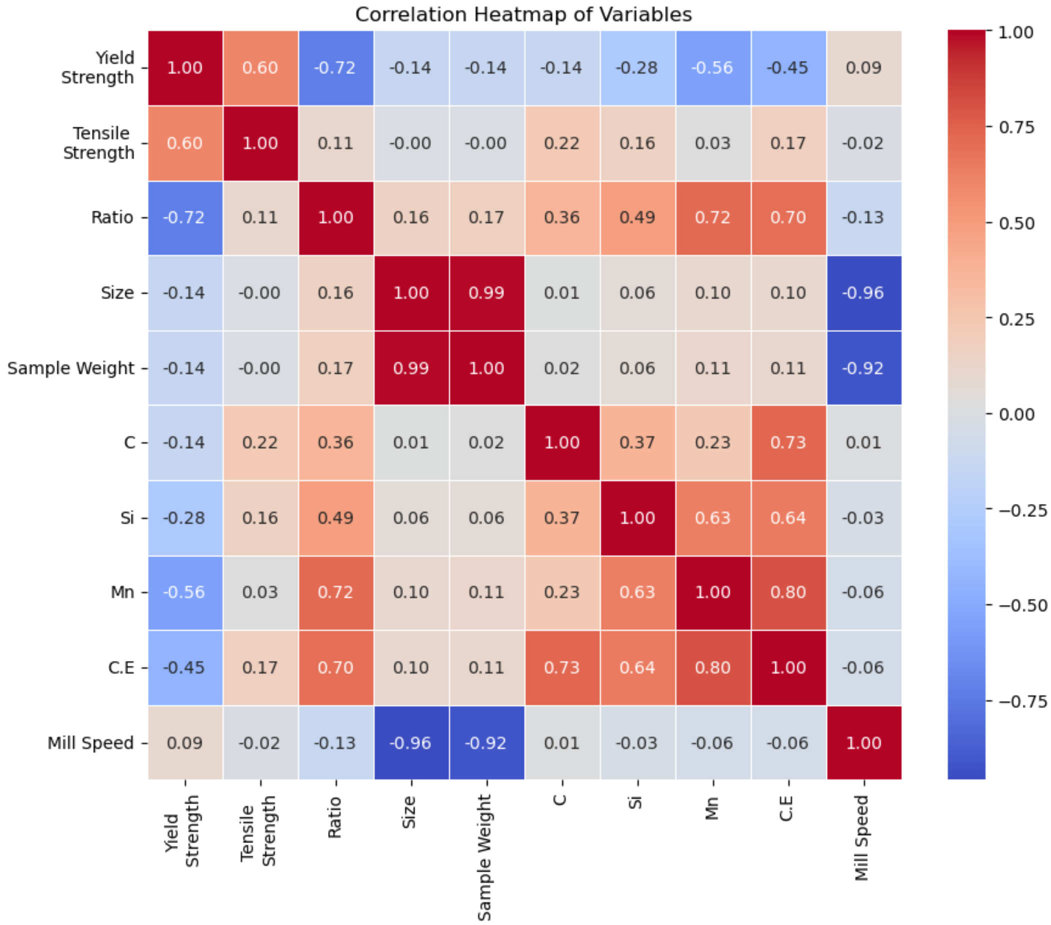

Figure 4: Heat map showing correlation between dependent and independent variables.

Practical Application in Rebar Rolling Campaigns:

In a pragmatic validation of our developed machine learning model, we executed a series of meticulously designed rebar rolling experiments across varying diameters and chemical compositions. The integration of critical factors such as mill speed, water flow rate, and precise chemical composition data derived from optical emission spectroscopy augmented the model's precision and predictive capabilities.

Examination and Significance:

The significant alignment was observed between predicted and actual yield strengths opens up an intriguing discussion about the practical applications and critical implications of our machine learning model. Its ability to accurately predict mechanical properties by amalgamating chemical composition and process data marks a significant milestone in data-driven decision-making within the steel industry.

Furthermore, the model's success in handling a myriad of input variables underscores its flexibility in capturing the intricate relationships within the rolling process. By leveraging these data-driven insights, manufacturers can fine-tune their rebar production processes, making them more precise and efficient. This infusion of predictive analytics into industrial operations has the potential to modernize conventional manufacturing practices, injecting data-driven decision-making into the core of production processes[18]. In essence, our research signifies a step towards a more efficient, sustainable, and data-driven future in construction materials manufacturing.

Conclusion: This study highlights the efficacy of a data-driven approach utilizing machine learning techniques for predicting mechanical properties in steel reinforcement bars (rebars).

Leveraging a meticulously curated dataset of over 10,300 samples and employing advanced machine learning algorithms integrated into the PyCaret model, accurate predictive models were developed. The approach focused on utilizing chemical composition and quenching parameters as essential input features, supported by thorough data preprocessing and cross-validation techniques to ensure robust model generalization and prevent overfitting.

The analysis demonstrates the model's capability to accurately predict Yield Strength (YS) and elucidates a positive linear relationship between actual and predicted values. Moreover, it also illustrates the model's effectiveness in predicting Ultimate Tensile Strength (UTS), with a slight dispersion of data points suggesting predictive accuracy across a range of values. It predicted the model's ability to predict the Ratio of UTS to YS (UTS/YS) with a relatively small average deviation between actual and predicted values and a high explanatory power.

The study provides valuable insights into the relationship between input variables and key mechanical properties, including Yield Strength (YS), Ultimate Tensile Strength (UTS), and the Ratio of UTS to YS (UTS/YS). These findings contribute significantly to the enhancement of quality control and optimization in the steel manufacturing industry, underscoring the potential of machine learning in addressing critical challenges in materials science and engineering.

Acknowledgement:

The authors would like to acknowledge indigenous steel industries for providing the real dataset for our preliminary work on the ML model. The authors also acknowledge that this manuscript has not been published before, nor is it under consideration by any research journal

Author's Contribution:

1) **Muhammad Ali Siddiqui (Supervisor)**: Conceptualization of the study, oversight of data collection, finalization of the manuscript, project supervision, and validation of predicted outcomes.
2) **Taimoor Hussain Qureshi**: Drafting of the manuscript, data cleaning, optimization, data analysis, and literature review.
3) **Syed Dabeer Mehdi Naqvi**: Implementation of the machine learning algorithm, creation of graphs and tables, data analysis, and interpretation.
4) **Muhammad Zeeshan**: Mathematical formulation of the machine learning linear regression model, theoretical analysis, and validation.
5) **Minhal Waseem**: Draft writing, formatting of the manuscript, verification of the machine learning algorithm, and literature review.
6) **Zain Noreen:** Formatting of the manuscript, verification of the machine learning algorithm, coordination of contributions, and final proofreading.

Conflict of interest:

There is no conflict of interest among authors for publishing this manuscript in IJIST.

REFERENCES

[1] ASTM A615/A615M-12, "Standard Specification for Deformed and Plain Carbon-Steel Bars for Concrete Reinforcement," Asm Int., 2012, doi: 10.1520/A0615.

[2] British Standards Institution., "BS 4449:2005+A2:2009 : steel for the reinforcement of concrete : weldable reinforcing steel : bar, coil and decoiled product : specification," vol. 3, p. 27, 2009.

[3] F. Djavanroodi and A. Salman, "Variability of Mechanical Properties and Weight for Reinforcing Bar Produced in Saudi Arabia," IOP Conf. Ser. Mater. Sci. Eng., vol. 230, no. 1, 2017, doi: 10.1088/1757-899X/230/1/012002.

[4] Z. Wang, P. Liu, Y. Xiao, X. Cui, Z. Hu, and L. Chen, "A Data-Driven Approach for Process Optimization of Metallic Additive Manufacturing under Uncertainty," J. Manuf. Sci. Eng. Trans. ASME, vol. 141, no. 8, 2019, doi: 10.1115/1.4043798.

[5] M. Mohtasham Moein et al., "Predictive models for concrete properties using machine learning and deep learning approaches: A review," J. Build. Eng., vol. 63, no. October 2022, 2023, doi: 10.1016/j.jobe.2022.105444.

[6] Nwakamma Ninduwezuor-Ehiobu et al., "Tracing the Evolution of Ai and Machine Learning Applications in Advancing Materials Discovery and Production Processes," Eng. Sci. Technol. J., vol. 4, no. 3, pp. 66–83, 2023, doi: 10.51594/estj.v4i3.552.

[7] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00592-x.

[8] D. M. Jones, J. Watton, and K. J. Brown, "Comparison of hot rolled steel mechanical property prediction models using linear multiple regression, non-linear multiple regression and non-linear artificial neural networks," Ironmak. Steelmak., vol. 32, no. 5, pp. 435–442, 2005, doi: 10.1179/174328105X48151.

[9] T. M. H. Hope, "Linear regression," in Machine Learning: Methods and Applications to Brain Disorders by Andrea Mechelli; Sandra Vieira, Elsevier Science, 2020, pp. 67–81.

[10] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," J. Appl. Sci. Technol. Trends, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/JASTT1457.

[11] S. Jahanian and M. Mosleh, "Mathematical modeling of phase transformation of steel during quenching," J. Mater. Eng. Perform., vol. 8, no. 1, pp. 75–82, 1999, doi: 10.1361/105994999770347197.

[12] A. Stoll and P. Benner, "Machine learning for material characterization with an application for predicting mechanical properties," GAMM Mitteilungen, vol. 44, no. 1, pp. 1–21, 2021, doi: 10.1002/gamm.202100003.

[13] E. Ruiz, D. Ferreño, M. Cuartas, A. López, V. Arroyo, and F. Gutiérrez-Solana, "Machine learning algorithms for the prediction of the strength of steel rods: an example of data-driven manufacturing in steelmaking," Int. J. Comput. Integr. Manuf., vol. 33, no. 9, pp. 880–894, 2020, doi: 10.1080/0951192X.2020.1803505.

[14] R. H. F. Murta, F. D. Braga, P. P. N. Maia, O. B. F. Diógenes, and E. P. de Moura, "Mathematical modelling for predicting mechanical properties in rebar manufacturing," Ironmak. Steelmak., vol. 48, no. 2, pp. 161–169, 2021, doi: 10.1080/03019233.2020.1749357.

[15] M. V. Carneiro, T. T. Salis, G. M. Almeida, and A. P. Braga, "Prediction of Mechanical Properties of Steel Tubes Using a Machine Learning Approach," J. Mater. Eng. Perform., vol. 30, no. 1, pp. 434–443, 2021, doi: 10.1007/s11665-020-05345-0.

[16] A. Mukhopadhyay and A. Iqbal, "Prediction of mechanical property of steel strips using multivariate adaptive regression splines," J. Appl. Stat., vol. 36, no. 1, pp. 1–9, 2009, doi: 10.1080/02664760802193252.

[17] C. Singh Tumrate, S. Roy Chowdhury, and D. Mishra, "Development of Regression Model to Predicting Yield Strength for Different Steel Grades," IOP Conf. Ser. Earth Environ. Sci., vol. 796, no. 1, 2021, doi: 10.1088/1755-1315/796/1/012033.

[18] J. Krumeich, S. Jacobi, D. Werth, and P. Loos, "Big data analytics for predictive manufacturing control - A case study from process industry," Proc. - 2014 IEEE Int. Congr. Big Data, BigData Congr. 2014, pp. 530–537, Sep. 2014, doi: 10.1109/BIGDATA.CONGRESS.2014.83.