

Received: 18.12.2023

Revised: 30.01.2024

Accepted: 16.02.2024



DOI: 10.17804/2410-9908.2024.1.018-027



MODELING THE PROOF STRENGTH OF PIPE STEELS TEMPERED WITH THE APPLICATION OF ARTIFICIAL INTELLIGENCE METHODS

M. F. Gafarov^{1, a), *} and K. Yu. Okishev^{2, b)}

¹South Ural State University (National Research University),
76 Lenina Ave., Chelyabinsk, 454080, Russia

²Ural Federal University named after the first President of Russia B. N. Yeltsin,
19 Mira St., Ekaterinburg, 620002, Russia

a)  <https://orcid.org/0000-0001-8479-9363>  gafarovmax@yandex.ru

b)  <https://orcid.org/0000-0002-6331-3050>  konstantin.okishev@urfu.ru

*Corresponding author. E-mail: gafarovmax@yandex.ru

Address for correspondence: pr-kt Lenina, 76, Chelyabinsk, 454080, Russia

The paper demonstrates the results of modeling the proof strength in pipe steels improved by tempering heat treatment. The main types of models used in this study are described, and information about the pros and cons of different approaches to modeling the target variable is summarized. Empirical equations relating hardness to yield strength and tensile strength are given. The role of the parameter n in these equations is indicated. The reasons for choosing the applied set of independent variables in the models are explained. The distribution of the target variable in the data sample is shown, and information about the feature space used for each of the models is provided. A general description of the source data is given. The structure of the main data sample is studied by the DBSCAN clustering method and the t-SNE dimension reduction algorithm. The reason for splitting the sample into clusters is substantiated in the context of reducing the spread of the predicted value of proof strength. The effectiveness of splitting the sample is estimated by using the measure of the spread of n . Various regression models for predicting yield strength are compared. It is shown that the regression model based on gradient boosting over the decision trees (LightGBM) has the smallest prediction error among the models considered. The permutation significance of the features of the model with the smallest prediction error is determined, the calculated significance of the features being compared with that from the metallurgical theory. The validity of the obtained prediction models is evaluated in view of the significance of the features and the metric estimate used in this study. The hypothesis of using a proxy variable (n) obtained from theoretical calculations as a predictor in the yield strength prediction model is tested. It is demonstrated that the application of the grouping method together with the parameter n makes it possible to obtain satisfactory prediction results on a smaller feature space.

Keywords: heat treatment of steels, machine learning, mathematical modeling, proof strength, pipe steels

References

1. Gulyaev, A.P. *Metallovedenie* [Physical Metallurgy]. Metallurgiya Publ., Moscow, 1986, 544 p. (In Russian).
2. Xie, Q., Suvarna, M., Li, J., Zhu, X., Cai, J., and Wang, X. Online prediction of mechanical properties of hot rolled steel plate using machine learning. *Materials & Design*, 2021, 197, 109201. DOI: 10.1016/j.matdes.2020.109201.

3. Bhattacharyya, T., Singh, S.B., Dey, S., Bhattacharyya, S., Bleck, W., and Bhattacharjee, D. Microstructural prediction through artificial neural network (ANN) for development of transformation induced plasticity (TRIP) aided steel. *Materials Science and Engineering: A*, 2013, 565 (10), 148–157. DOI: 10.1016/j.msea.2012.11.110.
4. Okishev, K.Yu. Calculation of TTT diagrams of isothermal austenite decomposition in structural steels. *Vestnik PNIPU. Mashinostroenie, Materialovedenie*, 2020, 22 (2), 82–89. DOI: 10.15593/2224-9877/2020.2.10. (In Russian).
5. Pavlina, E.J. and Van Tyne, C.J. Correlation of yield strength and tensile strength with hardness for steels. *Journal of Materials Engineering and Performance*, 2008, 17, 888–893. DOI: 10.1007/s11665-008-9225-5.
6. Lee, J.-Y., Kim, M., and Lee, Y.-K. Design of high strength medium-Mn steel using machine learning. *Materials Science and Engineering: A*, 2022, 843, 143148. DOI: 10.1016/j.msea.2022.143148.
7. Cui, C., Cao, G., Li, X., Gao, Z., Liu, J., and Liu, Z. A strategy combining machine learning and physical metallurgical principles to predict mechanical properties for hot rolled Ti micro-alloyed steels. *Journal of Materials Processing Technology*, 2023, 311, 117810. DOI: 10.1016/j.jmatprotec.2022.117810.
8. Gafarov, M., Okishev, K., and Makovetskiy, A. Predicting the hardness of pipe steels using machine learning methods. In: *2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2022, 1051–1056. DOI: 10.1109/ICIEAM54945.2022.9787169.
9. Okishev, K.Yu., Gafarov, M.F., Pavlova, K.P., Makoveckiy, A.N., and Gafarova, E.A. Construction and analysis of models for predictions of flow limits of pipe steel after thermal treatment using methods of machine learning. *Kuznechno-Shtampovochnoe Proizvodstvo. Obrabotka Materialov Davleniem*, 2022, 10, 10–17. (In Russian).
10. Available at: <https://scikit-learn.org/stable/modules/sgd.html#regression>
11. Cahoon, J.R., Broughton, W.H., and Kutzak, A.R. The determination of yield strength from hardness measurements. *Metallurgical and Materials Transactions*, 1971, 2, 1979–1983. DOI: 10.1007/BF02913433.
12. Van der Maaten, L.J.P. and Hinton, G.E. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 2008, 9, 2579–2605.
13. Schubert, E., Sander, J., Ester, M., and Kriegel, H.P. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 2017, 42 (3), 1–21. DOI: 10.1145/3068335.
14. Available at: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.iqr.html>
15. Available at: <https://lightgbm.readthedocs.io/en/stable/>
16. Available at: https://scikitlearn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html

Подана в журнал: 18.12.2023

УДК 669.017+51-74



DOI: 10.17804/2410-9908.2024.1.018-027



МОДЕЛИРОВАНИЕ УСЛОВНОГО ПРЕДЕЛА ТЕКУЧЕСТИ ТРУБНЫХ СТАЛЕЙ ПОСЛЕ УЛУЧШЕНИЯ С ПОМОЩЬЮ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

М. Ф. Гафаров^{1, а), *}, К. Ю. Окишев^{2, б)}

¹Федеральное государственное автономное образовательное учреждение высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)», пр-кт Ленина, 76, г. Челябинск, 454080, Россия

²Федеральное государственное автономное образовательное учреждение высшего образования «Уральский федеральный университет имени первого Президента России Б. Н. Ельцина», ул. Мира, 19, г. Екатеринбург, 620002, Россия

^{а)}  <https://orcid.org/0000-0001-8479-9363>  gafarovmax@yandex.ru;

^{б)}  <https://orcid.org/0000-0002-6331-3050>  konstantin.okishev@urfu.ru

*Ответственный автор. Электронная почта: gafarovmax@yandex.ru
Адрес для переписки: пр-кт Ленина, 76, г. Челябинск, 454080, Россия

В статье демонстрируются результаты моделирования условного предела текучести в трубных сталях после термообработки улучшением. Описываются основные типы моделей, используемые в работе, обобщается информация о плюсах и минусах разных подходов к моделированию целевой переменной. Приводятся эмпирические уравнения связи твердости, предела текучести и предела прочности. Указывается роль параметра n в приведенных уравнениях. Объясняются причины выбора применяемого набора независимых переменных в моделях. Показывается распределение целевой переменной в выборке данных, приводится информация о признаковом пространстве, использованном для каждой из рассмотренных моделей. Представлено общее описание исходных данных. Исследуется структура основной выборки данных с помощью метода кластеризации DBSCAN и алгоритма снижения размерности t-SNE. Обосновывается причина дробления выборки на кластеры в контексте снижения разброса прогнозируемой величины условного предела текучести. Оценивается эффективность разбиения выборки с помощью меры разброса введенного параметра n . Проводится сравнение различных регрессионных моделей прогнозирования предела текучести. Показывается, что регрессионная модель на основе градиентного бустинга над деревьями решений (LightGBM) имеет наименьшую ошибку прогнозирования среди рассмотренных моделей. Определяется перестановочная значимость признаков модели с наименьшей ошибкой прогнозирования, приводится сравнение вычисленной значимости признаков с данными металлургической теории. Оценивается валидность полученных моделей прогнозирования с учетом значимости признаков и метрической оценки, используемой в данной работе. Проверяется гипотеза об использовании прокси-переменной (параметра n), полученной на основе теоретических выкладок, в качестве предиктора модели предсказания предела текучести. Демонстрируется, что использование метода группировки совместно с параметром n позволяет получать удовлетворительные результаты прогнозирования на меньшем признаковом пространстве.

Ключевые слова: термообработка сталей, машинное обучение, математическое моделирование, условный предел текучести, трубные стали

1. Введение

Важным направлением исследований в металлургии является создание работоспособных моделей для прогнозирования результатов термической обработки сталей в различных условиях [1]. В качестве целевых переменных в разрабатываемых моделях, как правило, используют такие метрические показатели, как предел прочности, предел текучести, твердость, доли микроструктурных составляющих и т. д. [2, 3]. Конечная прогностическая модель термообработки с удовлетворительными величинами ошибки прогнозирования может быть использована как инструмент для решения производственных задач. К основным категориям решаемых задач могут относиться поиск причин брака продукции, оптимизация режимов термообработки, разработка новых систем легирования сталей и другие.

Получение конечных моделей прогнозирования при термической обработке сталей осложняется многими факторами и зависит от исследуемых групп сплавов. В данной работе рассматриваются марки доэвтектоидных низко- и среднелегированных сталей для моделирования пределов текучести после распространенного вида термической обработки – закалки с последующим высоким отпуском (улучшения). Сложность предсказания предела текучести (и других механических свойств) в таких сталях заключается в многообразии различных микроструктурных состояний с разными концентрациями легирующих элементов, которые могут изменяться при проведении термообработки [4]. Кроме того, для контроля и подробного описания происходящих изменений в процессе термической обработки требуется подбор достаточного количества предикторов (независимых признаков), что часто затруднено или невозможно в производственных условиях.

Для создания моделей в таких условиях используют различные методы математического моделирования. Одним из наиболее распространенных методов является построение линейных регрессионных моделей [4]. Отличительными особенностями таких моделей являются простота, хорошая интерпретируемость, а также небольшой объем выборки, требующийся для построения модели. Однако в случае наличия нелинейных взаимосвязей между предикторами и целевой переменной качество таких моделей будет определяться подготовкой признакового пространства.

Помимо классических линейных моделей, в ряде работ [2, 3, 6, 7] для получения низких значений ошибок прогнозирования целевой переменной после термообработки использовались модели, построенные с помощью методов искусственного интеллекта, а именно методов машинного обучения. Такие модели позволяют лучше обобщать закономерности на обучающей выборке, получать меньший разброс и дисперсию результатов прогнозирования и фактических значений. К недостаткам данного типа моделей можно отнести бóльший по сравнению с линейными моделями размер выборки, требующийся для получения адекватных результатов, а также низкую интерпретируемость.

В данной работе проводится сравнение классических линейных методов моделирования предела текучести и моделей, построенных с помощью методов машинного обучения.

2. Постановка задачи и методы решения

2.1. *Общее описание данных*

В качестве исходных данных были использованы экспериментальные данные термообработки (закалка и последующий высокий отпуск) образцов трубных сталей различного типоразмера. Выбор одного типа термообработки объясняется массовостью применения и единообразием конечных микроструктур после отпуска на фиксированных наборах марочно-го состава.

Независимые переменные для регрессионных моделей были условно разделены на категории. Выбор признакового пространства для моделей производился на основе опыта прошлых работ [8, 9] и экспертной оценки исследователей. Описание признакового пространства представлено в таблице 1.

Таблица 1

Описание признакового пространства

Категория признака	Признаки, входящие в категорию
Режимы термообработки	Температура закалки, температура отпуска, время отпуска, время выдержки в аустенитной области
Химический состав	Массовое содержание химических элементов: С, Mn, Si, Cr, Mo, Nb, V, Cu, Ni, Al, S, P, N, As, Ti, Ca, B, W, Pb, Sn, Sb
Типоразмер	Диаметр трубы, толщина стенки
Прочностные характеристики	Предел прочности

Помимо классического подхода к построению регрессионных уравнений в данной работе также анализируется метод построения моделей предела текучести с помощью расчетного параметра n .

Расчетный параметр n определялся по уравнениям (1, 2) с помощью градиентного спуска [10] с погрешностью до 10^{-6} :

$$TS = \left(\frac{H}{2,9}\right) \left(\frac{n}{0,217}\right)^n; \quad (1)$$

$$YS = \left(\frac{H}{3}\right) (0,1)^n. \quad (2)$$

Здесь TS – предел прочности, МПа; YS – предел текучести, МПа; H – твердость стали по шкале Виккерса, МПа. Данные уравнения использовались в аналогичных работах [5, 11] при моделировании предела текучести. Таким образом, величина n является переходным звеном в цепочке $TS \rightarrow n \rightarrow YS$. В настоящей работе, помимо построения классических моделей и моделей на основе методов машинного обучения, проверяется гипотеза о применимости параметра n в качестве основного предиктора в моделях предсказания предела текучести при разбиении выборки на кластеры. Предполагается, что параметр n , используемый таким образом, позволит снизить разброс значений прогнозируемой величины внутри кластеров.

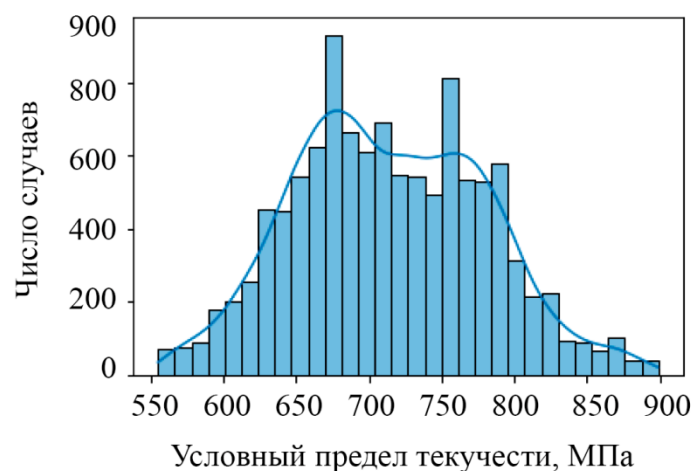


Рис. 1. Распределение условного предела текучести образцов в выборке

Объем результирующей выборки для построения и валидации моделей составил более 12 000 строк-векторов. При этом количество марок сталей составило 40 единиц. Марки стали, составляющие большую часть выборки, таковы: 30Г2, 18ХМФБ, 30ХГМА, 37Г2Ф, 38Г2С. Толщина стенки труб находилась в диапазоне 5–10 мм, диаметр труб 90–200 мм. Распределение целевого признака – предела текучести – по всей выборке представлено на рис. 1.

Распределение параметра n по маркам сталей вне зависимости от типоразмера труб представлено на рис. 2.

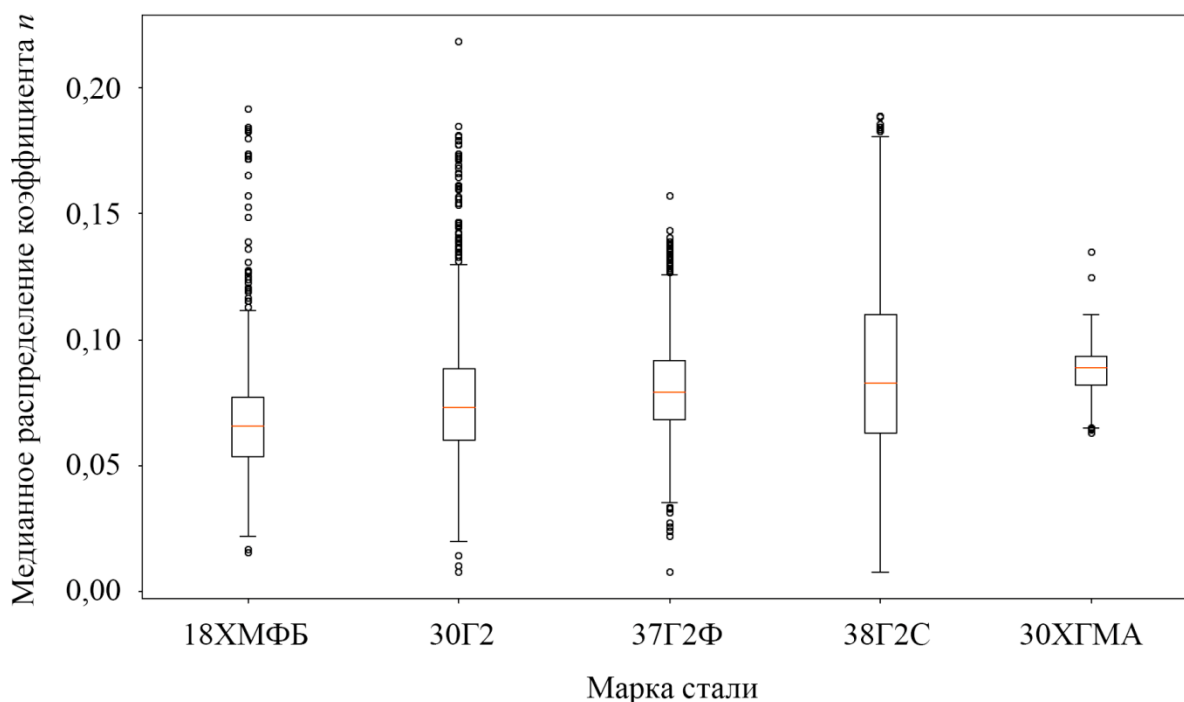


Рис. 2. Распределение коэффициента n образцов в выборке

Из анализа рис. 2 следует, что разброс значений и дисперсия внутри групп каждой марки стали без деления по типоразмеру велики. Это означает, что использование модели прогнозирования предела текучести, основанной только на уравнениях (1), (2), приведет к значительной дисперсии фактических и прогнозных значений.

2.2. Подготовка выборки для построения моделей методом группировки

Для того чтобы прогнозировать предел текучести сталей с помощью уравнений (1) и (2), требуется эффективно сгруппировать данные так, чтобы снизить дисперсию величины n , для чего независимые параметры (табл. 1) и расчетный параметр n использовались в качестве переменных для модели кластеризации. Первоначально размерность данных была понижена до 2-мерного пространства методом t-SNE [12], затем к полученному массиву был применен алгоритм кластеризации DBSCAN [13]. Внутри размеченных кластеров были определены средние коэффициенты n . Проекция массива точек со сниженной размерностью на плоскость представлена на рис. 3.

Из рис. 3 следует, что алгоритму DBSCAN удалось разделить различные точки массива со сниженной размерностью по кластерам. При этом видно, что кластеры имеют различные коэффициенты n . Детальный анализ кластеров показал, что принятие решения об отнесении того или иного образца выборки к выбранному кластеру осуществляется в основном за счет сравнения параметров «диаметр», «толщина стенки», «марка стали». Такое разделение

является логичным, поскольку при типичных режимах термообработки стандартных марок сталей комбинация этих параметров является определяющим признаком для свойств полученных труб. Важно понять, насколько разделение с помощью определенных трех параметров «диаметр», «толщина стенки», «марка стали» будет хуже, чем с использованием алгоритма DBSCAN. Для оценки меры разброса параметра n была использована величина межквартильного размаха (IQR) [14]. В табл. 2 приведено сравнение нескольких методов группировки данных для оценки средней величины разброса n .

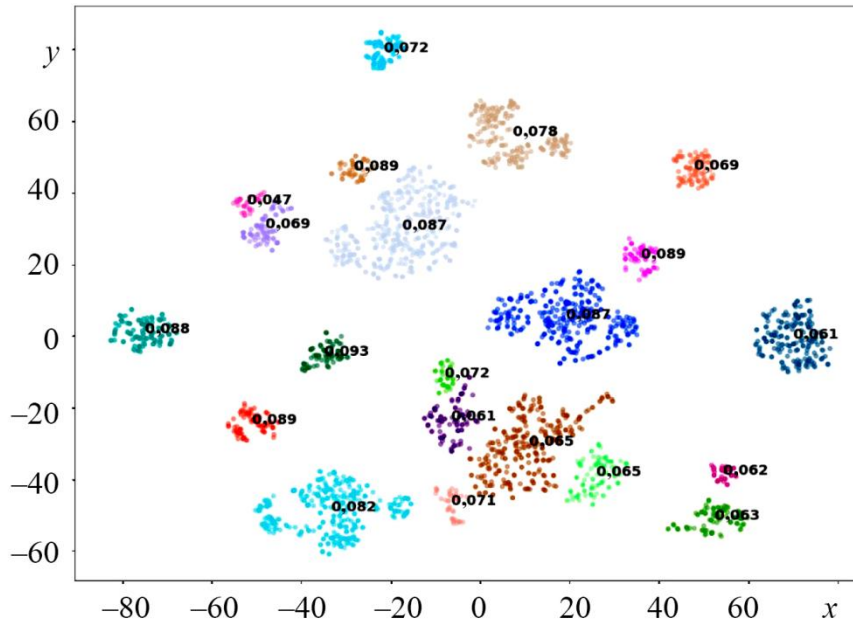


Рис. 3. Проекция точек на плоскость со значениями коэффициентов n по кластерам

Таблица 2

Сравнение методов группировки образцов стали

Метод группировки	Средняя величина IQR
По марке стали	0,026
По марке стали, толщине стенки, диаметру	0,022
DBSCAN на всем признаковом пространстве	0,02

В табл. 2 использовались группы, составленные только из марок сталей, перечисленных на рис. 2. Таким образом, группировка с использованием алгоритма DBSCAN позволяет снизить разброс значений n внутри групп, однако простая группировка по марке стали, толщине стенки и диаметру также обеспечивает сравнимое снижение разброса значений n . Для дальнейшего моделирования предела текучести был выбран подход группировки по марке стали, толщине стенки и диаметру.

2.3. Математические модели

В качестве основных регрессионных моделей для прогнозирования условного предела текучести были выбраны модель линейной регрессии (Linear), модель градиентного бустинга над решающими деревьями (LightGBM) [15] и модель группировки (Grouping). Выбор ли-

нейной модели и модели LightGBM обусловлен экспертной оценкой исследователей по опыту предыдущих исследований [8, 9].

Признаковым пространством для моделей линейной регрессии и LightGBM являются независимые переменные, описанные в табл. 1. На основе группировки независимых переменных «марка стали», «толщина стенки», «диаметр трубы» производится моделирование, при этом каждой комбинации этих признаков сопоставляется средняя величина n , определенная внутри группы. Далее определяется предел текучести с использованием величины n и уравнений (1), (2). Прогнозируемой величиной для каждой модели является условный предел текучести. В качестве основной метрики качества моделей использовалась средняя абсолютная ошибка (MAE) [16].

3. Результаты и обсуждение

Результаты оценки метрик полученных моделей на валидационной выборке представлены в табл. 3.

Таблица 3

Оценка ошибки прогнозирования моделей на валидационной выборке

Модель	MAE, МПа	R^2
Linear	26	0,65
Grouping	29	0,61
LightGBM	24	0,72

На рисунке 4 представлена диаграмма рассеяния фактических и прогнозных значений на валидационной выборке.

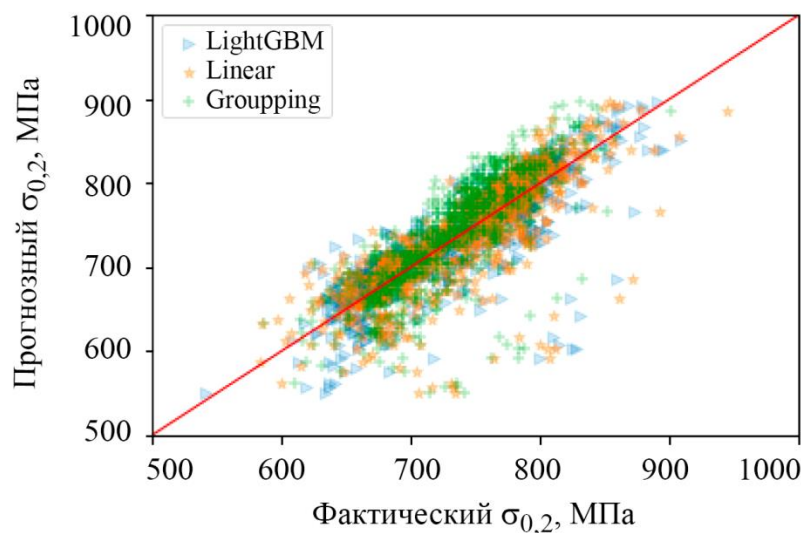


Рис. 4. Диаграмма рассеяния фактических и предсказанных значений предела текучести на валидационной выборке

Таким образом, лучшей моделью прогнозирования среди построенных является модель LightGBM, что подтверждает результаты прошлых работ [8, 9]. В то же время ошибки других моделей являются сопоставимыми с удовлетворительным результатом прогнозирования. При этом результаты моделирования предела текучести с использованием расчетного

параметра n и простого способа группировки демонстрируют хорошую сходимость фактических и прогнозных значений предела текучести.

Анализ значимости признаков для лучшей модели LightGBM показал, что, помимо таких существенных признаков, как *временное сопротивление*, *толщина стенки*, *диаметр*, для самых массовых марок сталей (рис. 2), наиболее значимыми признаками также являются *средняя температура отпуска* и *длительность отпуска*, что соответствует металлургической теории [1].

Несмотря на удовлетворительные оценки метрик, в некоторых случаях отмечаются большие отклонения между фактическими и прогнозными значениями (рис. 3). Такие значения можно отнести к категории выбросов в силу того, что в данной работе используется объемная выборка, в которой могут присутствовать, пусть и в небольшом процентном соотношении, как нарушения технологических режимов термообработки, так и ошибки в записи исходных данных. Дальнейшее повышение точности рассмотренных регрессионных моделей возможно путем добавления новых технических параметров (разбиения по зонам нагрева внутри печей, уточнение скорости нагрева и охлаждения труб и т. д.).

В целом можно утверждать, что разработанные модели могут быть применены на производстве для первичного контроля основных признаков с целью недопущения отклонения рассмотренных механических свойств от нормы. Кроме того, в случае получения бракованной продукции, построенные модели могут быть полезны при идентификации нарушений технологических режимов.

4. Заключение

В результате исследования были разработаны модели, позволяющие прогнозировать предел текучести сталей в низко- и среднелегированных трубных сталях после проведения термообработки, предусматривающей закалку и последующий высокий отпуск; проведен анализ моделей с целью сравнения точности прогнозирования различными подходами; выполнена валидация лучшей модели как с помощью метрических показателей, так и с помощью сравнения значимости признаков, при этом модельная значимость признаков соответствует металлургической теории. Показано, что регрессионные модели машинного обучения способны прогнозировать условный предел текучести с меньшей ошибкой, чем классический подход, основанный на линейных моделях. Продемонстрировано, что примененный метод группировки данных совместно с использованием параметра n обеспечивает получение стабильных оценок величин условного предела текучести.

Литература

1. Гуляев А. П. Металловедение : учебник для вузов. – 6-е изд., перераб. и доп. – М. : Металлургия, 1986. – 544 с.
2. Online prediction of mechanical properties of hot rolled steel plate using machine learning / Q. Xie, M. Suvarna, J. Li, X. Zhu, J. Cai, X. Wang // *Materials & Design*. – 2021. – Vol. 197. – P. 109201. – DOI: 10.1016/j.matdes.2020.109201.
3. Microstructural prediction through artificial neural network (ANN) for development of transformation induced plasticity (TRIP) aided steel / T. Bhattacharyya, S. B. Singh, S. Dey, S. Bhattacharyya, W. Bleck, D. Bhattacharjee // *Materials Science and Engineering: A*. – 2013. – Vol. 565. – P. 148–157. – DOI: 10.1016/j.msea.2012.11.110.
4. Окишев К. Ю. Расчет диаграмм изотермического распада аустенита в конструкционных сталях // *Вестник ПНИПУ. Машиностроение, материаловедение*. – 2020. – № 2. – С. 82–89. – DOI: 10.15593/2224-9877/2020.2.10.
5. Pavlina E. J., Van Tyne C. J. Correlation of yield strength and tensile strength with hardness for steels // *Journal of Materials Engineering and Performance*. – 2008. – Vol. 17. – P. 888–893. – DOI: 10.1007/s11665-008-9225-5.

6. Lee J.-Y., Kim M., Lee Y.-K. Design of high strength medium-Mn steel using machine learning // Materials Science and Engineering: A. – 2022. – Vol. 843. – P. 143148. – DOI: 10.1016/j.msea.2022.143148.
7. A strategy combining machine learning and physical metallurgical principles to predict mechanical properties for hot rolled Ti micro-alloyed steels / C. Cui, G. Cao, X. Li, Z. Gao, J. Liu, Z. Liu // Journal of Materials Processing Technology. – 2023. – Vol. 311. – P. 117810. – DOI: 10.1016/j.jmatprotec.2022.117810.
8. Gafarov M., Okishev K., Makovetskiy A. Predicting the hardness of pipe steels using machine learning methods // 2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – 2022. – P. 1051–1056. – DOI: 10.1109/ICIEAM54945.2022.9787169.
9. Построение и анализ моделей предсказания пределов текучести трубных сталей после термической обработки с помощью методов машинного обучения / К. Ю. Окишев, М. Ф. Гафаров, К. П. Павлова, А. Н. Маковецкий, Е. А. Гафарова // Кузнечно-штамповочное производство. Обработка материалов давлением. – 2022. – № 10. – С. 10–17.
10. URL: <https://scikit-learn.org/stable/modules/sgd.html#regression>
11. Cahoon J. R., Broughton W. H., Kutzak A. R. The Determination of yield strength from hardness measurements // Metallurgical and Materials Transactions. – 1971. – Vol. 2. – P. 1979–1983. – DOI: 10.1007/BF02913433.
12. Van der Maaten L. J. P., Hinton G. E. Visualizing high-dimensional data using t-SNE // Journal of Machine Learning Research. – 2008. – Vol. 9. – P. 2579–2605.
13. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN / E. Schubert, J. Sander, M. Ester, H. P. Kriegel // ACM Transactions on Database Systems (TODS). – 2017. – Vol. 42 (3). – P. 1–21. – DOI: 10.1145/3068335.
14. URL: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.iqr.html>
15. URL: <https://lightgbm.readthedocs.io/en/stable/>
16. URL: https://scikitlearn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.