Metals & corrosion

Microstructure–property relation and machine learning prediction of hole expansion capacity of high-strength steels

Wei Li[1,2,*], Martina Vittorietti[3], Geurt Jongbloed[3], and Jilt Sietsma[1]

[1] Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
[2] Materials Innovation Institute (M2i), Mekelweg 4, 2628 CD Delft, The Netherlands
[3] Department of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands

ABSTRACT

The relationship between microstructure features and mechanical properties plays an important role in the design of materials and improvement of properties. Hole expansion capacity plays a fundamental role in defining the formability of metal sheets. Due to the complexity of the experimental procedure of testing hole expansion capacity, where many influencing factors contribute to the resulting values, the relationship between microstructure features and hole expansion capacity and the complexity of this relation is not yet fully understood. In the present study, an experimental dataset containing the phase constituents of 55 microstructures as well as corresponding properties, such as hole expansion capacity and yield strength, is collected from the literature. Statistical analysis of these data is conducted with the focus on hole expansion capacity in relation to individual phases, combinations of phases and number of phases. In addition, different machine learning methods contribute to the prediction of hole expansion capacity based on both phase fractions and chemical content. Deep learning gives the best prediction accuracy of hole expansion capacity based on phase fractions and chemical composition. Meanwhile, the influence of different microstructure features on hole expansion capacity is revealed.

Address correspondence to E-mail: w.li-3@tudelft.nl

GRAPHICAL ABSTRACT
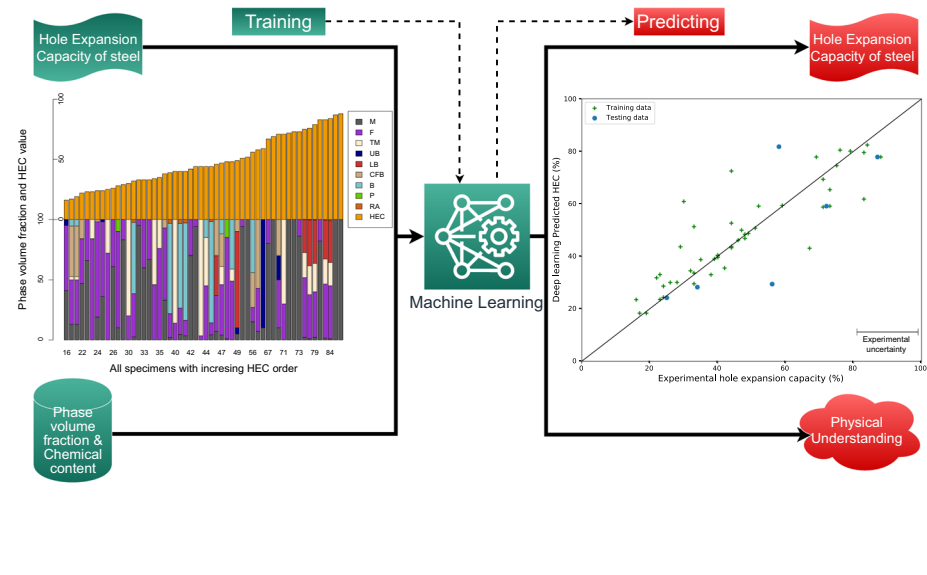
Introduction

Advanced high-strength steels are widely used in industrial applications. Besides high strength and good ductility, stretch-flangeability is an important mechanical property which controls the quality of shaping many metallic components. Hence, the hole expansion capacity (HEC), which describes the formability and edge cracking resistance of sheet metals, is one of the most important mechanical properties in for instance the automotive industry. Figure 1 shows the most common test procedure for the determination of HEC following standard ISO 16630 [1]. The sheet metal is first punched with an initial hole of diameter D_0 of 10 mm. The punched hole is then widened with a conical punch (60° angle) until the first through-thickness crack appears, at the final hole diameter D_h. The hole expansion capacity (λ) is then calculated with

$$\lambda = \frac{D_h - D_0}{D_0}. \tag{1}$$

The results are considered useful when the thickness of the sheet material is below 2.5 mm, even though the standard allows thicknesses up to 6 mm. Due to the complexity of determining the hole expansion capacity, many testing factors, such as punch edge quality [2–4] and crack determination [5], could influence the testing result.

The HEC is not yet well understood in terms of its relationship with the microstructure of the metal. Many studies have been performed on the relations between HEC and microstructure features, processing parameters and other mechanical properties, such as tensile strength and hardness [3, 5–8], but results either are not convincing due to the limited number of data or do not give an overall picture on the effects of multiple phases due to the specific materials choices. Meanwhile, recent progress in the field of HEC of multi-phase steels results in a better understanding of the relation between HEC and fracture toughness, which can be related to microstructure features through damage and fracture models [9, 10]. It has been shown that HEC is closely related to the capacity to resist the initiation of micro-cracks and their propagation [11–13]. The connection of fracture behavior and microstructure features and heterogeneities can then be extended to the understanding of the HEC behavior [14–17]. The study from de Geus et al. [15] shows that fracture initiation correlates strongly with the local microstructural morphology. Meanwhile, the laminography observations performed by Kahziz et al. [18] reveal the damage evolution on both the punched and machined edges,
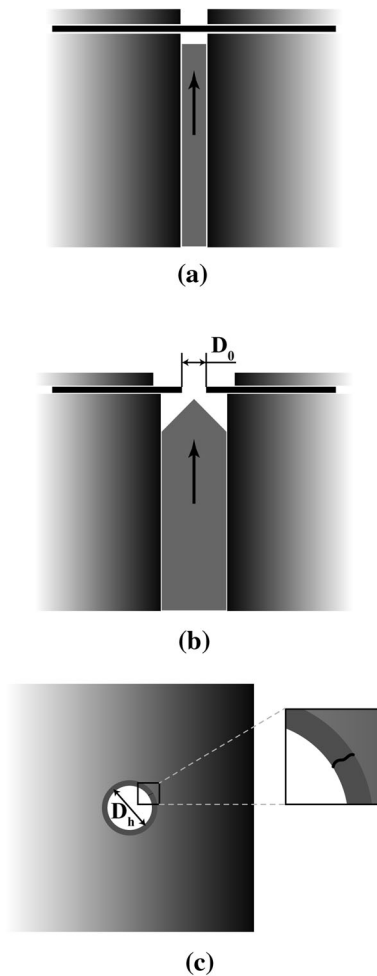
Figure 1 Schematic illustration of HEC test after ISO 16630. The standard prescribes $D_0 = 10$ mm.

which indicates the possibility of building predictive models based on physical understanding.

Table 1 summarizes the present interpretation in the literature on the influence of different phases on the HEC from various studies on multi-phase steels. Except for ferrite and possibly austenite, all phases are reported to have a negative effect on the hole expansion property. These effects are often explained by the hardness difference between the hard phases and the soft phases, but with no clear physical explanation [2, 6, 19].

The information obtained from Table 1 is rather limited and qualitative, since the trends of changing HEC on different phases are always gathered from a very limited number of data points (i.e., less than 10). Besides, the results shown in Table 1 only concern the relation of a single microstructure feature with hole expansion capacity. When dealing with complex-phase steels, the combined effect of the phases has not been studied yet.

In order to study the relations between hole expansion capacity and microstructure features in more detail, 55 groups of data containing the composition of phases and chemical content corresponding to the HEC values are collected, as shown in Table 4 (see Appendix A), from a final report of a research project of the Research Fund for Coal and Steel [20]. As the original report does not make full use of these data, it is valuable to have a deep look into these data and to derive more comprehensive understanding in addition to Table 1. In the present paper, HEC is fully investigated on its relation to phase fractions individually, to the combination of phases and to the number of phases. To quantify the relations, different statistical regression methods are applied to enable prediction of the HEC on the basis of both phase fractions and chemical content, while also giving the importance ranking of different microstructure features.

Data analysis

As shown in Table 4 (see Appendix A), the studied dataset contains 55 complex-phase steel specimens' results. For each specimen, the hole expansion capacity (HEC, %) with phase fractions in volume percentage and chemical composition in weight percentage are collected. The identified phases are martensite (M), ferrite (F), tempered martensite (TM), upper bainite (UB), lower bainite (LB), carbide-free bainite (CFB), bainite (B), pearlite (P) and retained austenite (RA). For these microstructures, ferrite, martensite and bainite can be present as matrix phases, while pearlite and retained austenite are always secondary phases. The average standard deviation of measuring HEC three times on the same steel grade is $\pm 9\%$, which is calculated from the work by Chen et al. [22]. Meanwhile, there is also research showing a standard deviation of 15% on HEC values for martensitic steels [23].

HEC in relation to volume fraction of phases

Volume fraction of individual phases

Based on the obtained data, the individual influence of phase volume fractions on hole expansion capacity

is shown in the scatter plots in Fig. 2. Figure 2a and b show the scatter plot of martensite (without tempered martensite) and ferrite fraction in relation to hole expansion capacity, while Fig. 2c and d show the total bainite (the sum of upper bainite, lower bainite, carbide-free bainite and bainite) and retained austenite volume fractions in relation to hole expansion capacity. The straight line in Fig. 2b is a linear fitting of all data points of ferrite volume fraction and hole expansion capacity. All curved lines in Fig. 2 are based on the *scatter.smooth* function in R [24], which uses the *loess* (local polynomial regression fitting) function [25]. The lines are merely a guide to the eye for the main trends.

As shown in Fig. 2a, there is a clear valley in the plot of the relation between HEC and martensite volume fraction, which indicates either low martensite volume fraction (lower than 20%) or high martensite volume fraction (higher than 80%) tends to have the possibility to reach relatively high HEC. Meanwhile, HEC is always low when the martensite volume fraction falls between 20% and 70%. For ferrite in Fig. 2b, the relation is not as clear as for martensite, but a very distinct observation is that only low HEC values are found above 50%. When ferrite volume fraction is lower than 50%, there is no clear relation between HEC and ferrite volume fraction. Low HEC values occur in the region where ferrite volume fraction is higher than 50%, with only one exception: No. 23 in Table 4, that consists of a large volume fraction of ferrite and secondary phase pearlite. An opposite trend to martensite is shown in

Fig. 2c when looking into the relation between the total bainite volume fraction and HEC. High HEC values are found only between 30% and 40% bainite. Figure 2d shows the relation between HEC and the secondary phase retained austenite. There are obviously two stages in the relation of HEC with retained austenite volume fraction in Fig. 2d. The lower volume fraction of retained austenite shows higher HEC than the group of higher volume fraction. In Fig. 2d, the bainite fractions are also indicated for the microstructures. Relating these values to Fig. 2c, the relation between HEC and bainite fraction, it shows that the microstructures with low RA fractions all lie in the optimum range of bainite fraction. The values of HEC for zero retained austenite fraction, with the average on the green line, lie within the shaded area in Fig. 2d, at the level of the values for 2–4% RA. The present data therefore do not give a conclusive view on the influence of retained austenite on HEC.

The dataset is unfortunately very limited on pearlite. Only three microstructures contain pearlite, of which one is the exceptional No. 23. The other two are No. 26 (15% P, 84% F, 1% M, $\lambda = 48\%$) and No. 29 (10% P, 80% F, 10% M, $\lambda = 28\%$). The difference between these two HEC values is therefore primarily the result of the difference in martensite and pearlite fractions. The reduction from $\lambda = 48\%$ for 1% martensite to $\lambda = 28\%$ for 10% martensite is stronger than the general trend in Fig. 2a, which points at a positive effect of pearlite on the HEC.

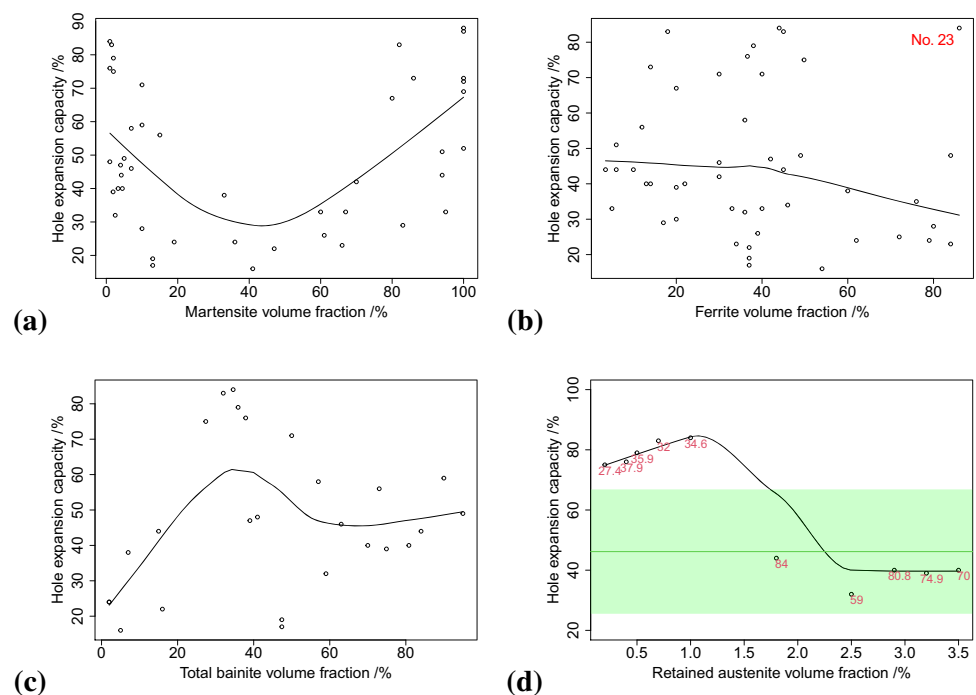**Table 1** Reported influence of different phases on HEC for several multiphase steels

Features	Influence on the HEC	
Ferrite	Positive	Polygonal ferrite is especially good. Higher strength of the ferritic matrix is positive to HEC. [3, 7, 19]
Martensite	Negative	Higher martensite volume fraction lowers the HEC. HEC decreases from 30% to 15%, when martensite volume fraction increases from 16% to 18%. [7, 19]
Bainite	Negative	Better than martensite and pearlite, but still negative; more bainite gives rise to a lower HEC. [6, 7]
Grain boundary cementite, pearlite	Negative	Microcracks often observed at the interface of elongated pearlite or grain boundary cementite. [6]
Retained austenite	Ambiguous	The stable or carbon-enriched retained austenite films enhance the HEC due to the reduction of the surface damage on hole-punching and the promotion of the TRIP effect on hole expansion. [21]
Voids and other impurities	Negative	Voids and other irregularities found on the edge surfaces lower the HEC. [3]

Difference between volume fraction of phases

Many researchers proposed that the HEC is closely related to the difference in mechanical behavior between hard and soft phases [2, 6, 7, 19]. Here, we assume that the ferrite and retained austenite are soft phases, while martensite, bainite and pearlite are hard phases. The relation between HEC and the volume fraction difference is shown in Fig. 3a. The scatter plot shows an increase of HEC when the hard phase volume fraction is increasing. When the dataset is divided into two groups, as the boxplot in Fig. 3b shows, the microstructures in the group with more than 50% volume fraction of hard phase have significantly higher HEC than the group with more than 50% volume fraction of soft phase. This indicates that HEC displays a relation with the strength of materials. The lack of high HEC values for microstructures with a higher fraction of soft phases coincides with the observation in Fig. 2b.

The one exception mentioned in "Volume fraction of individual phases" section, No. 23 in Table 4, is also marked in Fig. 3. It clearly shows that this No. 23 sample is an outlier with exceptionally high HEC while containing more soft phase, which is considered to be an artifact of the testing procedure. Hence, in the following statistical analysis, this No. 23 sample is deleted from the dataset.

HEC in relation to combinations of phases

As discussed in the previous section, certain phases (ferrite and martensite) have a distinct impact on HEC. The phase compositions with the increasing order of HEC are plotted in Fig. 4 with both combined and non-combined fractions of similar phases (applied for martensite and for bainite). Considering samples which have relatively high HEC, two kinds of phase composition are occurring frequently, either fully or nearly fully martensite, or a combination of ferrite, martensite and bainite with the volume ratio around 2:1:1. This indicates the significant contribution of martensite and bainite to HEC. It is also found that most two-phase martensite/ferrite microstructures, especially with a high ferrite fraction, have low HEC values.

HEC in relation to number of phases

Figure 5 shows the boxplot of HEC in relation to number of phases. In order to quantify this relationship, a linear regression model with the number of phases as explanatory variable and the HEC as the response variable is introduced. The number of phases is recognized as a factor, i.e., categorical variables. The number of phases has either 5 or 4 categories, corresponding to the non-combined phase fractions or the condition of combining all bainite and

Figure 2 HEC relation with martensite (**a**), ferrite (**b**), total bainite (**c**) and retained austenite (**d**) volume fractions. The numbers in (**d**) given with the data points are the total bainite fraction in percent. The green line shows the average HEC value, while the green shaded region shows the standard deviation, of the microstructures without RA.

all martensite. In Table 2, the t-value is the estimate (second column, the coefficient for each input variable) divided by its standard error (third column). By comparing this t-value to the *Student's t distribution*, the p-value can be calculated [26]. A small p-value (typically below 0.05) indicates that there is a relation between the explanatory variable and the response variable. The intercept of the linear model is at the one-phase category, which indicates that the one-phase category is set as the baseline. The model shows the change of HEC values of increasing number of phases based on the one-phase category.

Figure 5 shows that the one-phase category has the highest HEC values, while in Table 2, the p-values (last column) for one, two, three and four phases in non-combed phase fractions (a) and one, two and three phases in combining all bainite and all martensite conditions (b) are all below 0.05, which indicates that it is highly unlikely that the coefficient is equal to zero instead of the current value of the estimate [27, 28]. Since the values of the estimate are all negative except for the one-phase condition, it indicates that the one-phase category has the highest HEC. In this dataset, only the pure martensite structure appears in the one-phase category; hence, the result suggests that for pure martensite structure, the hole expansion capacity is significantly higher with respect to HEC values for microstructures with two, three or four phases, as shown in Fig. 5 and Table 2. Only the five phases without combining and four phases with combining have increased HEC, since these structures belong to the ones mentioned in "HEC in relation to combinations of phases" section which have the combination of ferrite, martensite and bainite with the volume ratio around 2:1:1. These microstructures all have a low volume fraction of retained austenite, and both martensite and tempered martensite are present.
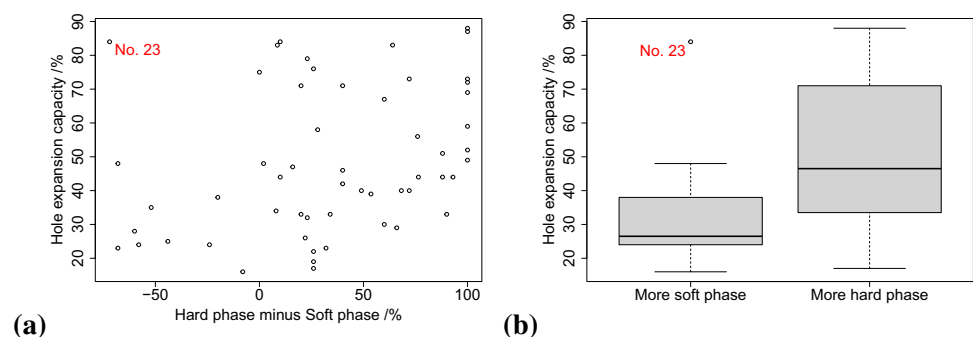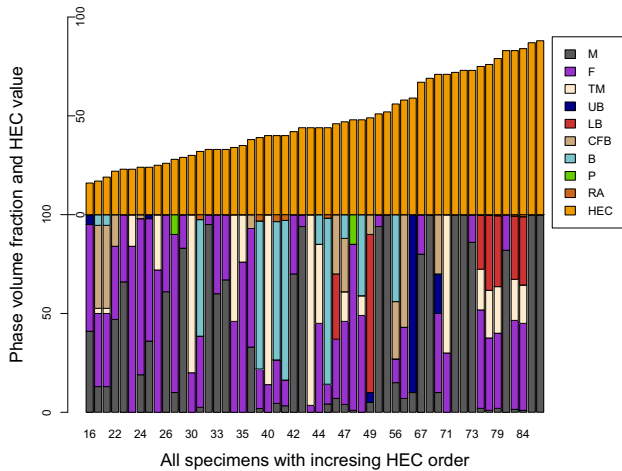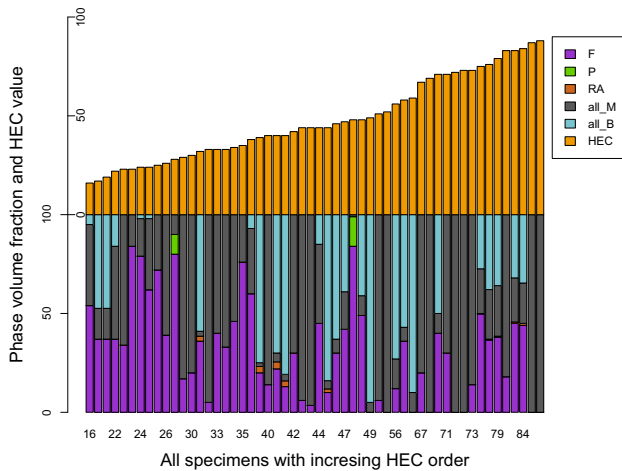
LASSO selection of importance phases

Because of the large number of phases and a single target variable, hole expansion capacity, a statistical method called least absolute shrinkage and selection operator (LASSO) is employed as described comprehensively in previous work [29] and in Appendix B.1.

The LASSO regression is performed on only the matrix phases, i.e., martensite, ferrite and bainite. In order to avoid the collinearity, the samples with only ferrite and martensite phases are excluded in this regression. Collinearity is a condition where two or more independent variables are highly correlated, which tends to inflate the coefficient for one variable and hence leads to wrong estimates of the coefficients [26]. In Fig. 6, with the decrease of the LASSO penalty parameter $\log(\lambda_e)$ [29], more input variables (phases) are included in the linear regression. The first four phases showing up in Fig. 6 from the high-λ_e side of the graph are lower bainite, martensite, upper bainite and ferrite. Since in the LASSO analysis just a linear function between HEC and the phase volume fractions is adopted, LASSO is not sufficient to fully explain the relationships, but LASSO does give an indication of certain phases which make the most significant contribution to the influence on HEC, namely lower bainite, martensite, upper bainite and ferrite. Meanwhile, LASSO shows that lower bainite and upper bainite have a clear positive effect on hole expansion capacity and martensite has a negative effect. Here, the negative effect from martensite seems to be different from the trend shown in Fig. 2a. This is because the samples with only martensite and ferrite have more than 50% of martensite, which are not included in the LASSO regression. Hence, the negative effect of martensite from LASSO only shows the effect for 0–50% martensite, which is therefore the same as the trend shown in Fig. 2a.
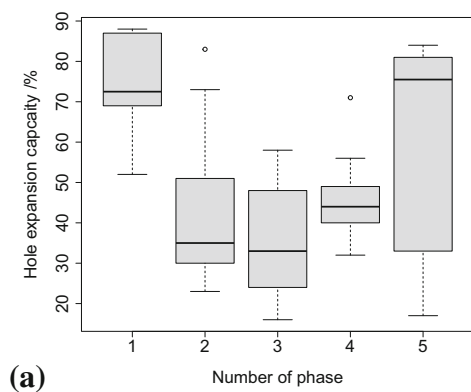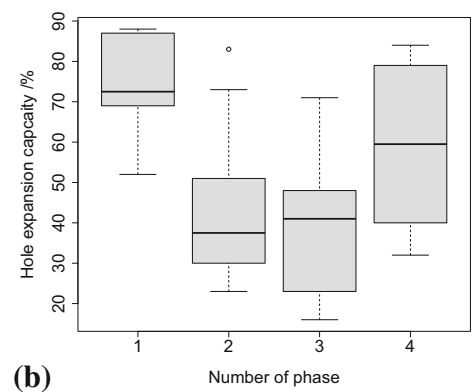
Figure 3 HEC relation with difference between hard and soft phase fractions (**a**) and HEC difference for two groups with respect to the volume fraction of hard and soft phase (**b**).

(a)

(b)

(a)

(b)

Figure 4 Phase composition with increasing HEC order, **a**: all the individual phases are present, **b**: combining all the martensite together and all the bainite together.

Table 2 Summary of results for linear regression between number of phases and HEC for both non-combing phase (**a**) and combining all bainite and all martensite (**b**), corresponding to Fig. 5

(a) Number of phases	Estimate	Std. error	t value	P(>|t|)
1 (Intercept)	74	7.3	10.04	0.0000
2	− 31	8.3	− 3.70	0.0005
3	− 39	9.3	− 4.16	0.0001
4	− 27	9.5	− 2.87	0.0060
5	− 14	9.7	− 1.39	0.1698
(b) Number of phases	Estimate	Std. error	t value	P(>|t|)
1 (Intercept)	74	7.3	10.14	0.0000
2	− 30	8.2	− 3.72	0.0005
3	− 36	8.5	− 4.19	0.0001
4	− 14	9.2	− 1.56	0.1253

Prediction of HEC with both phase fraction and chemical contents

Machine learning has been widely adopted in various applications in materials science due to its powerful data processing and high prediction performance [30–36]. In order to predict the HEC with both phase fractions and chemical content based on the data gathered in Table 4, we selected five different machine learning methods:

1. Linear regression (lm),
2. Linear regression with Elastic Net regularization (glmnet),

Figure 5 HEC relation with number of phases (**a**) and with number of phases while combining all bainite and all martensite (**b**).

(a)

(b)

3. Conditional inference tree regression (ctree2),
4. Random forest regression (cforest),
5. Deep learning (keras).

Detailed information on these methods is given in Appendix B. The first four methods are applied using the 'caret' library [37], adopted in the *R* [24] environment. For the first four methods, tenfold cross-validation is repeated five times. There is no tuning parameter in lm. For glmnet, the tuning grid for mixing percentage α is ten grids from 0 to 1 and 50 grids from 0.0001 to 50 for regularization parameter λ. For ctree2, the tuning grid for max tree depth *maxdepth* is five grids from 1 to 5 and ten grids from 0 to 1 for (1 minus p-value) threshold *mincriterion*. For cforest, the tuning grid for randomly selected predictors *mtry* is 15 grids from 1 to 15. Deep learning is applied using the 'keras' library [38], which uses TensorFlow [39] as backend in python. The network consists of two hidden layers. Both hidden layers are dense layers with 100 and 50 neurons, respectively. Both hidden layers use the activation function *relu* [40]. The model compiles with optimizer *Adam* [41]. The training epoch is 600 with batch size of 32 and validation split of 5%. The modeling process follows a route consisting of five steps:

1. data partitioning into training and testing set (random: 90% of the data in the training set and 10% in the testing set);
2. feed training data to train the model,
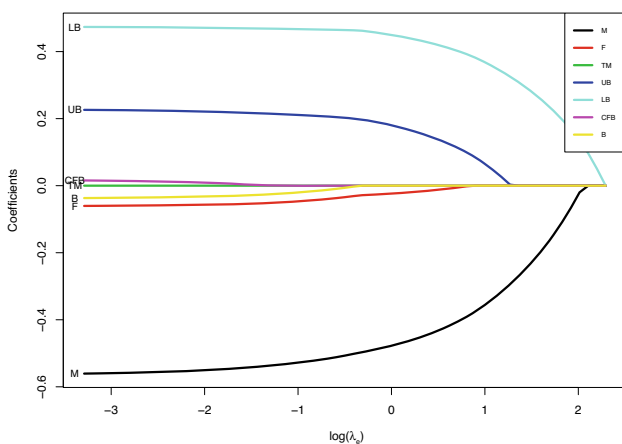3. predict testing target (HEC) using the trained model;

4. calculate the performance (calculate RMSE on both training and testing data);
5. repeat steps 1–4 ten times (tenfold cross-validation) and calculate the mean performance, i.e., the average RMSE on both training and testing data over ten repeated runs.

In step 4, the RMSE is calculated on both training dataset and testing dataset based on the predicted hole expansion capacities $\lambda_{p,i}$, the real hole expansion capacities $\lambda_{r,i}$ and number of samples N in the dataset as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\lambda_{p,i} - \lambda_{r,i})^2}{N}}. \tag{2}$$

Machine learning model performance

The performance of all five machine learning models is shown in Fig. 7. The two linear regression methods (1 and 2) and the conditional inference tree regression (3) clearly perform the worst with a high RMSE (root mean square error) on the testing dataset. The deep learning model shows the best performance with the lowest RMSE. The HEC prediction accuracy of the deep learning model is ±16%. Comparing the hole expansion testing error range of the experimental data acquired by Chen et al. [22], where the average standard deviation of testing three times on the same steel grade is ±9%, and the 15% standard deviation of experimental HEC values for martensitic steels [23], due to various testing conditions, such as edge surface quality and first crack determination timing, it can be concluded that deep learning predictions reach a similar degree of accuracy as experiments, where the 9% accuracy for the training dataset indicates an experimental accuracy of that magnitude. In Fig. 8, the deep learning-predicted HEC is plotted against the experimental HEC, with the experimental test error shown in the bottom-right corner. It can be seen that based on the learning from the training data points, deep learning can give confident prediction of the testing data points. With the improvement in the experimental data quality and increase in quantity of the data, the authors believe that the prediction accuracy can be further enhanced.

Machine learning model interpretation

The conditional inference tree regression model and random forest regression model both give rise to a

Figure 6 LASSO plot on the influence of different phase fractions on HEC.

ranking of importance of the independent variables, which is shown in Fig. 9. The feature importance based on the conditional inference tree is calculated by the sum of the reduction of variance to the parent node weighted by the probability of reaching that node caused by the certain feature. A higher value indicates high importance. Random forest averages the importance of each feature from each tree to obtain the rank of importance of all features.

Ferrite, martensite and lower bainite are the three most important phases that affect the HEC, while Mn and Cr are the two chemical elements that most strongly affect the HEC. The possible reason why Mn and Cr are the most important chemical elements is that both have a positive effect on the phase fraction of lower bainite (with a Pearson correlation coefficient of 0.31). The Pearson correlation coefficient can theoretically range from -1 to 1. A value of 1 means total positive linear correlation, a value of -1 means total negative linear correlation, and a value of 0

Figure 10 The conditional inference tree regression model plot, phase fractions in percentage and chemical content in weight percent. Colors indicate the magnitude of the average HEC in the node.

Table 3 Summary of the conditional inference tree based on the range of HEC values

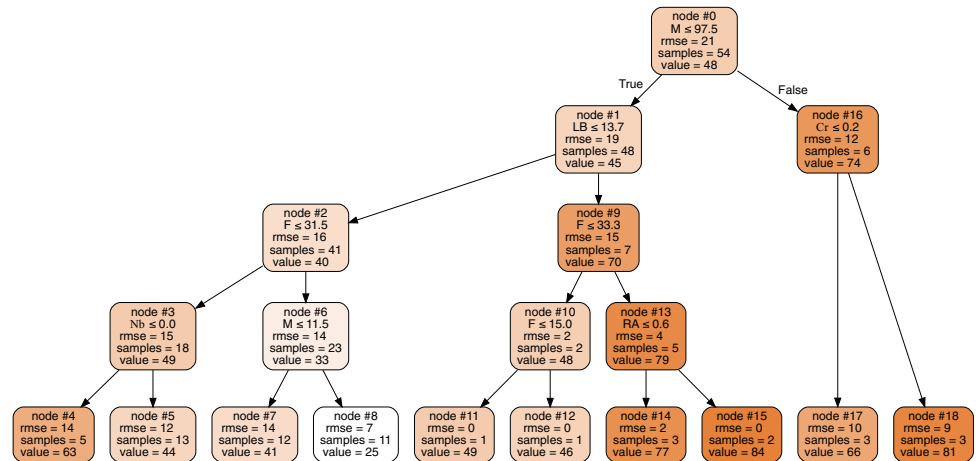| No.	HEC (%)	Phase fraction (%) and chemical content (wt%)	Node
I	25	$11.5 < f_M \leqslant 97.5$; $f_{LB} \leqslant 13.7$; $f_F > 31.5$	Node 8
II	40–50	$f_M \leqslant 97.5$; $f_{LB} \leqslant 13.7$; $f_F \leqslant 31.5$; $c_{Nb} > 0$	Node 5
		$f_M \leqslant 11.5$; $f_{LB} \leqslant 13.7$; $f_F > 31.5$	Node 7
		$f_M \leqslant 97.5$; $f_{LB} > 13.7$; $f_F \leqslant 33.3$	Node 10
III	60−70	$f_M \leqslant 97.5$; $f_{LB} \leqslant 13.7$; $f_F \leqslant 31.5$; $c_{Nb} = 0$	Node 14
		$f_M > 97.5$; $c_{Cr} \leqslant 0.2$	Node 17
IV	Over 70	$f_M \leqslant 97.5$; $f_{LB} > 13.7$; $f_F > 33.3$	Node 13
		$f_M > 97.5$; $c_{Cr} > 0.2$	Node 18

means no linear correlation [27, 42]. The effect of Mn and Cr is also shown by Guo et al. [43] who state that Mn improves strength to certain extent, while Cr improves ductility of bainitic steels. Higher ductility has positive effect on hole expansion capacity [44], while higher strength leads to lower hole expansion capacity [22].

With the conditional inference tree regression model, a decision tree can be built as shown in Fig. 10. At each node of the decision tree, one specific input variable is selected, according to algorithms mentioned in Appendix B.2, to separate the dataset into two subsets. For each node, the separation criterion, the root mean square error of samples in the node, the number of samples and the mean HEC value of all samples in the node are shown in the node box. The left arrow from the node box indicates the condition for separation is true, while the right arrow indicates it is false. Node 0 contains all 54 samples; its criterion is a martensite phase fraction smaller than 97.5%. This criterion is true for 48 samples with an average HEC of 45%, as shown in node 1, and it is false for six samples with an average HEC of 74%, as shown in node 16. The samples of each of

these nodes are further separated on the basis of subsequent criteria. The color of the node indicates its average HEC values.

From the decision tree, the trend of the influence of different independent variables is evidenced. It shows that the change in HEC with different variables is not monotonic. Table 3 summarizes the information from the decision tree based on the range of the HEC values corresponding to the phase fractions and chemical contents. Node 13 and node 18 in Fig. 10 classify the highest HEC with either fully martensitic structure or the combination of martensite, lower bainite and ferrite. Meanwhile, node 8 in Fig. 10 classifies the lowest HEC with more than 31.5% ferrite, less than or equal to 13.7% lower bainite and a martensite volume fraction between 11.5% and 97.5%.

Discussion

Comparing the summary in Table 1 and the statistic analysis in "HEC in relation to volume fraction of phases" section, there are some clearly contradictory

results. The summary in Table 1 shows that only ferrite has a positive effect on hole expansion capacity, while all other phases have a negative effect. But the analysis from a large number of data, as presented in the present paper, shows more complicated effects due to varying volume fractions of different phases, other than simply positive or negative. This is mainly due to the limitation of the range of data in studies in Table 1. Most of the studies only observe a certain fraction range of certain phases, which is not representing the effect on HEC across the whole volume fraction range. The effect of phase fractions on HEC is complicated and cannot be expressed by simple monotonic functions.

Taking into account the analysis in "Data analysis" section, in whichever way the data are looked at, the most important phases which contribute to HEC are ferrite, martensite and lower bainite. Considering that many studies relate the HEC to the difference in hard/soft phases, these three phases actually take the most important role in hard/soft phases in steels, especially ferrite and martensite, which are most commonly seen the softest phase and the hardest phase. Statistics show that the higher the fraction of the hard phase is, the higher the HEC is. This reflects that the HEC is a strength-related mechanical property. The HEC value shows a valley at the intermediate volume fraction of martensite which is possibly related to the minimum fracture stain in dual-phase steels with the similar condition of martensite [45, 46]. This can be explained by damage nucleation and crack growth mechanics being favored by strength mismatch and the related increase in the local stress triaxiality. Meanwhile, certain combinations of phases also give high HEC, such as the combination of ferrite, martensite and bainite with the volume ratio around 2:1:1. This high HEC can be accounted for by the accommodation of stress by this specific volume combination of hard and soft phases, where the hard phase gives the overall strength and soft phase gives ductility for expansion under stress without cracking. But the ferrite/martensite combinations do not perform very well. Although with the analysis in this paper, the complicated relations between HEC and microstructure features are clearly shown, it is not possible to give a simple relation. However, with the help of deep learning, a reliable prediction (with an accuracy of ±16% on HEC, which is similar to the experimental accuracy) can be made with the combination of the volume fraction of each

phase and chemical content. Still, the accuracy of the prediction model highly depends on the amount of the data gathered and the accuracy of the data. Even though the dataset used in this study is a large dataset in the context of materials science, it is definitely limited and small in the field of so-called big data and traditional machine learning. Nevertheless, the present study shows that meaningful results can also be achieved with limited datasets. The authors believe that significant improvement in the prediction model can be made if the data will be enhanced, both in the amount and in the quality.

In this study, since the obtained dataset only contains the phase volume fractions and the chemical composition, the data analysis and prediction of HEC are only based on these two microstructure features. Even without considering many other microstructure features, such as grain size distribution, texture and grain morphology, which are normally considered to have distinct impact on mechanical behavior, this study shows valuable results with limited materials information.

Conclusions

This study focuses on data acquired from the literature to investigate the relation of phase volume fractions and chemical compositions with hole expansion capacity. The findings in this paper can guide some new physical investigations to unravel the root causes of the HEC behavior and consequently to the development of better steels. The following conclusions are drawn based on the analysis from different perspectives.

– The effect of phase fractions on HEC is complicated and cannot be expressed by simple monotonic functions. For martensite, volume fractions between 20% and 70% will lead to a low HEC. HEC slightly decreases with an increasing volume fraction of ferrite. Around 30% bainite gives a high HEC.
– Certain phases make significant contribution to the HEC, most prominently, ferrite, martensite and lower bainite.
– The higher the volume fraction of harder phases is, the higher the HEC is.
– Purely martensitic microstructure or microstructure with lower bainite tends to have higher HEC compared to other combinations of phases. High

HEC can also be achieved with the combination of ferrite, martensite and bainite with the volume ratio around 2:1:1.

– The applied deep learning model has better performance (with the prediction accuracy of $\pm 16\%$ on HEC) over the linear regression models and tree regression models on the prediction of HEC based on phase fraction and chemical content.

Acknowledgements

Declarations

Appendix

Raw data from literature

The collected and organized data from the work of Ryde et al. [20] about the hole expansion capacity with chemical composition and phase fractions are shown in Table 4.

Principles of regression methods

Linear regression with Elastic Net regularization (glmnet)

The Elastic Net regression of the target mechanical property f at the point $\mathbf{x}$ is

$$\hat{f}_{\text{ElasticNet}}(\mathbf{x}) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_j, \tag{3}$$

where x_j is the jth variable in the prediction point $\mathbf{x}$ and p is the number of independent variables. The estimate $\hat{\beta}_j$ is the corresponding coefficient in the Elastic Net which minimizes the objective function:

$$L(\beta_0, \beta_1, \ldots, \beta_p) = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2$$
$$+ \lambda_e \sum_{j=1}^{p}(\alpha|\beta_j| + (1-\alpha)\beta_j^2). \tag{4}$$

Here, n is the number of the data points, x_{ij} is the ith observation corresponding to the jth variable and y_i is the target mechanical property corresponding to the data point $\mathbf{x}_i$. Different from the LASSO method which is used in the previous work [29], here the shrinkage penalty has two parts [47], namely LASSO penalty (magnitude α) and Ridge penalty (magnitude $1 - \alpha$). The LASSO penalty is indifferent while solving the problem among a set of strong but correlated variables. The Ridge penalty, on the other hand, tends to shrink the coefficients of correlated variables toward each other. The Elastic Net penalty is a combination of the two, also a compromise [48]. The two regularization parameters (α and λ_e) are optimized within a certain tuning grid during the training process.

Conditional inference tree regression (ctree2)

A decision tree is a model in the form of a tree structure, which breaks the dataset into smaller and smaller subsets; hence, the tree structure is built up. In order to build a tree structure, the most important two main steps are needed: to choose the feature and to find the condition to split, i.e., the partitioning algorithm. The most popular implementations of the recursive partitioning criteria, such as 'CART' [49] and 'C4.5' [50], have the problem of overfitting and a

Table 4 Raw data from the literature [20] with hole expansion capacity (HEC, %) with phase fractions in percentage and chemical content in wt%, the phases are martensite, ferrite, tempered martensite, upper bainite, lower bainite, carbide-free bainite, bainite, pearlite and retained austenite.

No.	HEC	M	F	TM	UB	LB	CFB	B	P	RA	C	Mn	Si	Cr	Mo	Nb
1	52	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
2	67	80.0	20.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
3	88	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
4	73	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
5	51	94.0	6.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
6	87	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
7	72	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
8	33	95.0	5.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
9	69	100.0	0.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
10	42	70.0	30.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
11	33	60.0	40.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
12	83	82.0	18.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
13	33	67.0	33.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
14	26	61.0	39.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
15	73	86.0	14.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
16	29	83.0	17.0	0.0	0	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
17	23	66.0	34.0	0.0	0	0.0	0	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
18	44	94.0	6.0	0.0	0	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
19	38	33.0	60.0	0.0	0	0.0	7	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
20	24	19.0	79.0	0.0	0	0.0	2	0.0	0	0.0	0.135	1.50	0.50	0.0	0.00	0.015
21	22	47.0	37.0	0.0	0	0.0	16	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
22	24	36.0	62.0	0.0	2	0.0	0	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
23	84	0.0	86.0	0.0	0	0.0	0	0.0	14	0.0	0.135	1.50	0.50	0.0	0.00	0.015
24	16	41.0	54.0	0.0	5	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
25	71	10.0	40.0	0.0	20	0.0	30	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
26	48	1.0	84.0	0.0	0	0.0	0	0.0	15	0.0	0.135	1.50	0.50	0.0	0.00	0.015
27	59	10.0	0.0	0.0	90	0.0	0	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
28	46	7.0	30.0	0.0	0	33.0	30	0.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
29	28	10.0	80.0	0.0	0	0.0	0	0.0	10	0.0	0.135	1.50	0.50	0.0	0.00	0.015
30	49	5.0	0.0	0.0	5	80.0	10	0.0	0	0.0	0.140	1.70	0.18	0.3	0.15	0.000
31	19	13.0	37.0	2.6	0	0.0	42	5.4	0	0.0	0.170	1.70	0.18	0.3	0.15	0.000
32	17	13.0	37.0	2.6	0	0.0	42	5.4	0	0.0	0.170	1.70	0.18	0.3	0.15	0.000
33	56	15.0	12.0	0.0	0	0.0	29	44.0	0	0.0	0.170	1.70	0.18	0.3	0.16	0.000
34	47	4.0	42.0	15.0	0	0.0	27	12.0	0	0.0	0.110	2.10	0.15	0.0	0.20	0.020
35	58	7.0	36.0	0.0	0	0.0	57	0.0	0	0.0	0.140	2.10	0.20	0.3	0.00	0.020
36	35	0.0	76.0	24.0	0	0.0	0	0.0	0	0.0	0.144	1.52	0.48	0.0	0.00	0.018
37	30	0.0	20.0	80.0	0	0.0	0	0.0	0	0.0	0.172	1.59	0.51	0.0	0.00	0.017
38	71	0.0	30.0	70.0	0	0.0	0	0.0	0	0.0	0.170	1.63	0.47	0.0	0.00	0.017
39	44	0.0	3.5	96.5	0	0.0	0	0.0	0	0.0	0.178	1.41	0.44	0.0	0.00	0.016
40	40	0.0	14.0	86.0	0	0.0	0	0.0	0	0.0	0.135	1.53	0.18	0.0	0.00	0.015
41	25	0.0	72.0	28.0	0	0.0	0	0.0	0	0.0	0.125	1.52	0.20	0.0	0.00	0.015
42	34	0.0	46.0	54.0	0	0.0	0	0.0	0	0.0	0.125	1.52	0.20	0.0	0.00	0.015
43	23	0.0	84.0	16.0	0	0.0	0	0.0	0	0.0	0.125	1.52	0.20	0.0	0.00	0.015
44	44	0.0	45.0	40.0	0	0.0	0	15.0	0	0.0	0.130	1.00	0.20	0.0	0.00	0.015
45	48	0.0	49.0	10.0	0	0.0	0	41.0	0	0.0	0.130	1.00	0.20	0.0	0.00	0.015
46	75	2.0	49.8	20.6	0	27.4	0	0.0	0	0.2	0.140	2.10	0.20	0.3	0.00	0.020
47	84	1.0	44.0	19.4	0	34.6	0	0.0	0	1.0	0.140	2.10	0.20	0.3	0.00	0.020
48	83	1.5	45.0	20.8	0	32.0	0	0.0	0	0.7	0.140	2.10	0.20	0.3	0.00	0.020
49	76	1.0	36.6	24.1	0	37.9	0	0.0	0	0.4	0.140	2.10	0.20	0.3	0.00	0.020

Table 4 continued

No.	HEC	M	F	TM	UB	LB	CFB	B	P	RA	C	Mn	Si	Cr	Mo	Nb
50	79	2.0	38.0	23.6	0	35.9	0	0.0	0	0.5	0.140	2.10	0.20	0.3	0.00	0.020
51	32	2.5	36.0	0.0	0	0.0	0	59.0	0	2.5	0.170	1.70	0.18	0.3	0.15	0.015
52	40	4.5	22.0	0.0	0	0.0	0	70.0	0	3.5	0.170	1.70	0.18	0.3	0.15	0.015
53	44	4.2	10.0	0.0	0	0.0	0	84.0	0	1.8	0.170	1.70	0.18	0.3	0.15	0.015
54	40	3.3	13.0	0.0	0	0.0	0	80.8	0	2.9	0.170	1.70	0.18	0.3	0.15	0.015
55	39	1.9	20.0	0.0	0	0.0	0	74.9	0	3.2	0.170	1.70	0.18	0.3	0.15	0.015

selection bias toward covariates with many possible splits [51]. Therefore, the conditional inference tree regression, also known as unbiased recursive partitioning, was introduced [51]. Unlike selecting the variable and deciding split criteria based on Gini Impurity [49] or Information Gain [50], it uses a significance test procedure, e.g., permutation tests [52]. Conditional inference tree is proved to be well suited for both explanation and prediction.

Random forest regression (cforest)

A random forest is a meta-estimator (i.e., it combines the result of multiple predictions) which aggregates many decision trees. It is a bagging technique, i.e., bootstrap aggregation, which is done with random sampling with replacement and aggregation of the outputs at the end without preference to any model. Therefore, the cforest model used in this paper is the conditional random forest which can be simply seen as averaging multiple conditional inference tree results [48, 51, 53].

Deep learning (keras)

Deep Learning refers to deep neural networks. It is an artificial intelligence function involving multiple units, called neurons, which are connected to each other like a Web, to make the data processing in a nonlinear approach. The 15 input neurons build up the input layer, while the output layer, in this study, is just one neuron, i.e., hole expansion capacity. The hidden layers are in the middle. The kind of fully connected neural network is called multilayer perceptron. Data flow from the input layer through the hidden layers and finally arrive output layer. The mathematics for calculating the value Y of each neuron from the neurons in the previous layer is [48, 54]:

$$Y = F\left(b + \sum_{i=1}^{n} w_i x_i\right), \tag{5}$$

where w_i is the weight for the neuron with value x_i, n is the number of neurons and b is the bias of each layer, which is a constant for each layer. F is the activation function, and *relu* is used in this study, which adds complexity and dimensionality to the neuron network. While the network is trained by feeding it with input data, the weight and bias will be learned to correct themselves to minimize the loss function by the technique called back propagation. The loss function, in this case, is the mean squared error between the network calculated HEC and the real HEC corresponding to the input microstructure.

References

[1] ISO 16630:2017, (2017) Metallic materials - sheet and strip - hole expanding test. Standard, International Organization for Standardization

[2] Karelova A, Krempaszky C, Werner E, Tsipouridis P, Hebesberger T, Pichler A (2009) Hole Expansion of Dual-phase and Complex-phase AHS Steels - Effect of Edge Conditions. Steel Res Int 80(1):71–77. https://doi.org/10.2374/SRI08SP110

[3] Ishimaru E, Takahashi A, Ono N (2010) Effect of material properties and forming conditions on formability of high-purity ferritic stainless steel. Nippon Steel Tech Rep 99:26–32

[4] Xu L, Barlat F, Lee MG, Choi KS, Sun X (2012) Hole expansion of dual phase steels. WIT Trans Built Environ 124:75–83. https://doi.org/10.2495/HPSM120071

[5] Comstock R, Scherrer D, Adamczyk R (2006) Hole expansion in a variety of sheet steels. J Mater Eng Perform 15(6):675–683. https://doi.org/10.1361/105994906X150830

[6] Cho YR, Chung JH, Ku HH, Kim IB (1999) Effect of controlled cooling on the formability of TS 590 MPa grade hot-rolled high strength steels. Met Mater 5(6):571–578. https://doi.org/10.1007/BF03026307

[7] Fang X, Fan Z, Ralph B, Evans P, Underhill R (2003a) The relationships between tensile properties and hole expansion property of C-Mn steels. J Mater Sci 38(18):3877–3882. https://doi.org/10.1023/A:1025913123832

[8] Fang X, Fan Z, Ralph B, Evans P, Underhill R (2003b) Effects of tempering temperature on tensile and hole expansion properties of a C-Mn steel. J Mater Process Technol 132(1–3):215–218. https://doi.org/10.1016/S0924-0136(02)00923-8

[9] Casellas D, Lara A, Frómeta D, Gutiérrez D, Molas S, Pérez L, Rehrl J, Suppan C (2017) Fracture Toughness to Understand Stretch-Flangeability and Edge Cracking Resistance in AHSS. Metall Mater Trans A 48(1):86–94. https://doi.org/10.1007/s11661-016-3815-x

[10] Lara A, Frómeta D, Molas S, Rehrl J, Suppan C, Casellas D (2016) Relation between stretch-flangeability and fracture toughness in advanced high strength steels. Proceedings of the IDDRG2016, Linz, Austria

[11] Frómeta D, Tedesco M, Calvo J, Lara A, Molas S, Casellas D (2017) Assessing edge cracking resistance in AHSS automotive parts by the Essential Work of Fracture methodology. Journal of Physics: Conference Series 896(1):012102. https://doi.org/10.1088/1742-6596/896/1/012102

[12] Frómeta D, Parareda S, Lara A, Molas S, Casellas D, Jonsén P, Calvo J (2020) Identification of fracture toughness parameters to understand the fracture resistance of advanced high strength sheet steels. Eng Fract Mech 229(2019):106949. https://doi.org/10.1016/j.engfracmech.2020.106949

[13] Ismail K, Perlade A, Jacques PJ, Pardoen T (2021) Outstanding cracking resistance of fibrous dual phase steels. Acta Materialia 207:116700. https://doi.org/10.1016/j.actamat.2021.116700

[14] Bareggi A, Maire E, Bouaziz O, Di Michiel M (2012) Damage in dual phase steels and its constituents studied by X-ray tomography. Int J Fract 174(2):217–227. https://doi.org/10.1007/s10704-012-9692-4

[15] de Geus T, van Duuren J, Peerlings R, Geers M (2016) Fracture initiation in multi-phase materials: A statistical characterization of microstructural damage sites. Mater Sci Eng, A 673:551–556. https://doi.org/10.1016/j.msea.2016.06.082

[16] Lai Q, Bouaziz O, Gouné M, Brassart L, Verdier M, Parry G, Perlade A, Bréchet Y, Pardoen T (2015a) Damage and fracture of dual-phase steels: Influence of martensite volume fraction. Mater Sci Eng, A 646:322–331

[17] Lai Q, Bouaziz O, Gouné M, Perlade A, Bréchet Y, Pardoen T (2015b) Microstructure refinement of dual-phase steels with 3.5 wt% mn: influence on plastic and fracture behavior. Mater Sci Eng, A 638:78–89

[18] Kahziz M, Morgeneyer TF, Mazière M, Helfen L, Bouaziz O, Maire E (2016) In situ 3D synchrotron laminography assessment of edge fracture in dual-phase steels: quantitative and numerical analysis. Exp Mech 56(2):177–195. https://doi.org/10.1007/s11340-015-0076-3

[19] Hasegawa K, Kawamura K, Urabe T, Hosoya Y (2004) Effects of microstructure on stretch-flange-formability of 980 MPa grade cold-rolled ultra high strength steel sheets. ISIJ Int 44(3):603–609. https://doi.org/10.2355/isijinternational.44.603

[20] Ryde L, Lyytinen O, Peura P, Titova M, Vilander Granbom Y, Hebesberger T (2012) Cold-rolled complex-phase (CP) steel grades with optimised bendability, stretch-flangeability and anisotropy (CP-Steels). Publications Office of the European Union. RFCS final report, Luxembourg

[21] Ki Sugimoto, Sakaguchi J, Iida T, Kashima T (2000) Stretch-flangeability of a high-strength TRIP type bainitic sheet steel. ISIJ Int 40(9):920–926. https://doi.org/10.2355/isijinternational.40.920

[22] Chen X, Jiang H, Cui Z, Lian C, Lu C (2014) Hole expansion characteristics of ultra high strength steels. Procedia Eng 81:718–723. https://doi.org/10.1016/j.proeng.2014.10.066

[23] Atzema E, Borsutzki M, Braun M, Brockmann S, Buelter M, Carlsson B, Larour P, Richter A (2012) A European round robin test for the hole expansion test according to ISO 16630. In: Proceedings of the International Conference: New Developments in Sheet Metal Forming, Fellbach, Germany, pp 171–184

[24] R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

[25] Cleveland W, Grosse E, Shyu W (1992) Local regression models. In: JM Chambers and TJ Hastie (ed) Statistical models in S, Wadsworth & Brooks/Cole, chap 8

[26] James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning, Springer Texts in Statistics, vol 103. Springer, New York, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

[27] Heumann C, Schomaker M, Shalabh, (2016) Introduction to statistics and data analysis. Springer International Publishing, Cham https://doi.org/10.1007/978-3-319-46162-5

[28] Ghatak A (2017) Machine learning with R. Springer Singapore, Singapore https://doi.org/10.1007/978-981-10-6808-9

[29] Li W, Vittorietti M, Jongbloed G, Sietsma J (2020) The combined influence of grain size distribution and dislocation

density on hardness of interstitial free steel. J Mater Sci Technol 45:35–43. https://doi.org/10.1016/j.jmst.2019.11.025

[30] Wei J, Chu X, Sun X, Xu K, Deng H, Chen J, Wei Z, Lei M (2019) Machine learning in materials science. InfoMat 1(3):338–358. https://doi.org/10.1002/inf2.12028

[31] Santak P, Conduit G (2019) Predicting physical properties of alkanes with neural networks. Fluid Phase Equilibria 501:112259. https://doi.org/10.1016/j.fluid.2019.112259

[32] Jose R, Ramakrishna S (2018) Materials 4.0: Materials big data enabled materials discovery. Appl Mater Today 10:127–132. https://doi.org/10.1016/j.apmt.2017.12.015

[33] Bessa MA, Glowacki P, Houlder M (2019) Bayesian machine learning in metamaterial design: Fragile becomes supercompressible. Advanced Materials p 1904845, https://doi.org/10.1002/adma.201904845

[34] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, Zeller M, Friedler SA, Schrier J, Norquist AJ (2016) Machine-learning-assisted materials discovery using failed experiments. Nature 533(7601):73–76. https://doi.org/10.1038/nature17439

[35] Dimiduk DM, Holm EA, Niezgoda SR (2018) Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering. Integr Mater Manuf Innovation 7(3):157–172. https://doi.org/10.1007/s40192-018-0117-8

[36] Holm EA (2019) In defense of the black box. Science 364(6435):26–27. https://doi.org/10.1126/science.aax0162

[37] Kuhn M (2008) Building predictive models in r using the caret package. J Stat Softw 28(5):1–26

[38] Chollet F, et al. (2015) Keras. https://keras.io

[39] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/, software available from tensorflow.org

[40] Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv preprint arXiv:180308375

[41] Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980

[42] Hogg RV, Craig AT, McKean JW (2014) Introduction to mathematical statistics. Pearson Education, Boston, Mass

[43] Guo H, Zhou P, Am Zhao, Zhi C, Ding R, Jx Wang (2017) Effects of Mn and Cr contents on microstructures and mechanical properties of low temperature bainitic steel. J Iron Steel Res Int 24(3):290–295. https://doi.org/10.1016/S1006-706X(17)30042-0

[44] Ki Sugimoto, Kanda A, Kikuchi R, Si Hashimoto, Kashima T, Ikeda S (2002) Ductility and formability of newly developed high strength low alloy TRIP-aided sheet steels with annealed martensite matrix. ISIJ Int 42(8):910–915. https://doi.org/10.2355/isijinternational.42.910

[45] Landron C, Bouaziz O, Maire E, Adrien J (2010) Characterization and modeling of void nucleation by interface decohesion in dual phase steels. Scripta Mater 63(10):973–976. https://doi.org/10.1016/j.scriptamat.2010.07.021

[46] Allain S, Bouaziz O (2008) Microstructure based modeling for the mechanical behavior of ferrite-pearlite steels suitable to capture isotropic and kinematic hardening. Mater Sci Eng: A 496(1–2):329–336. https://doi.org/10.1016/j.msea.2008.06.009

[47] Zou H, Hastie T (2005) Addendum: Regularization and variable selection via the elastic net. J R Stat Soc: Ser B (Stat Methodol) 67(5):768–768. https://doi.org/10.1111/j.1467-9868.2005.00527.x

[48] Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer Series in Statistics, Springer, New York, New York, NY,. https://doi.org/10.1007/978-0-387-84858-7

[49] Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman and Hall/CRC

[50] Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann Publishers, San Mateo, Calif

[51] Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: A conditional inference framework. J Comput Graph Stat 15(3):651–674. https://doi.org/10.1198/106186006X133933

[52] Strasser H, Strasser H, Weber C (1999) On the asymptotic theory of permutation statistics. Mathematical Methods of Statistics 2

[53] Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8(1):25. https://doi.org/10.1186/1471-2105-8-25

[54] Haykin S, Haykin SS (2009) Neural networks and learning machines. Prentice Hall

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer