A data-driven and machine-learning study on microstructure-property relations in steel

Li, W.

DOI
10.4233/uuid:d5c26a98-199f-43d6-8997-3e38dd5d4bd5

Publication date
2022

Document Version
Final published version

Citation (APA)
Li, W. (2022). *A data-driven and machine-learning study on microstructure-property relations in steel*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:d5c26a98-199f-43d6-8997-3e38dd5d4bd5

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

A data-driven and machine-learning study on microstructure-property relations in steel

Wei LI

A data-driven and machine-learning study on microstructure-property relations in steel

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Monday 11 July 2022 at 12:30 hours

by

Wei LI

Master of Science in Metallurgical Engineering,
RWTH Aachen University, Aachen, Germany,
born in Nanjing, China.

This dissertation has been approved by the promotors:

Prof. dr. ir. J. Sietsma
Prof. dr. ir. G. Jongbloed

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Promotors:	
Prof. dr. ir. J. Sietsma,	Delft University of Technology
Prof. dr. ir. G. Jongbloed,	Delft University of Technology
Independent members:	
Prof. dr. ir. L.A.I. Kestens	Delft University of Technology
Prof. dr. A.J. Cabo	Delft University of Technology
Prof. dr.-Ing. U. Prahl	Technische Universität Bergakademie Freiberg
Prof. dr. D.J. Jensen	Technical University of Denmark
Dr. P.J.J. Kok	Ghent University

The research described in this thesis was carried out in the Department of Materials Science and Engineering and the Delft Institute of Applied Mathematics of Delft University of Technology, the Netherlands.

This research was carried out under project number S41.5.14547a in the framework of the Partnership Program of the Materials innovation institute M2i (www.m2i.nl) and the Technology Foundation TTW, which is part of the Netherlands Organization for Scientific Research (www.nwo.nl).

Printed by: ProefschriftMaken (www.proefschriftmaken.nl)

Front & Back: Wei Li

ISBN 978-94-6384-351-5

An electronic version of this dissertation is available at
https://repository.tudelft.nl.

Not everything that can be counted counts,
and not everything that counts can be counted.

William Bruce Cameron

Contents

List of Tables **xi**

List of Figures **xiii**

1 Introduction **1**
References 5

2 Microstructure to Mechanical properties **7**
2.1 Introduction 8
2.2 Phase/Precipitation related 8
2.2.1 Phase characteristics 8
2.2.2 Phase fraction influence 9
2.2.3 Precipitation influence 13
2.3 Grain related 15
2.3.1 Grain characteristics 15
2.3.2 Grain size influence 16
2.3.3 Grain size distribution influence 18
2.3.4 Texture influence 18
2.4 Dislocation related 19
2.4.1 Dislocation structure 19
2.4.2 Dislocation density 20
2.4.3 Dislocation density influence 22
References 26

I Microstructure to Deformation **31**

3 Hardness **33**
3.1 Introduction 35
3.2 Material and Methods 37
3.2.1 Material and processing 37
3.2.2 Grain size distribution determination 37
3.2.3 Dislocation density determination by X-Ray diffraction (XRD) 38
3.2.4 Hardness measurement 39
3.2.5 Variable selection method: LASSO 39
3.3 Results and discussion 40
3.3.1 Microstructure and hardness 40
3.3.2 Grain size distribution 40
3.3.3 Dislocation density 42

3.3.4 Correlation of hardness with grain size distribution and dislocation density . . . 42
3.4 Conclusions . . . 47
References . . . 48
Appendix . . . 51
3.A Descriptive statistics and histogram of all samples . . . 51

II Microstructure to Fracture **53**

4 Hole expansion capacity **55**
4.1 Introduction . . . 57
4.2 Data analysis . . . 60
4.2.1 HEC in relation to volume fraction of phases . . . 60
4.2.2 HEC in relation to combinations of phases . . . 62
4.2.3 HEC in relation to number of phases . . . 64
4.2.4 LASSO selection of importance phases . . . 65
4.3 Prediction of HEC with both phase fraction and chemical contents . . . 67
4.3.1 Machine learning model performance . . . 68
4.3.2 Machine learning model interpretation . . . 68
4.4 Discussion . . . 73
4.5 Conclusions . . . 74
References . . . 74
Appendix . . . 79
4.A Raw data from literature . . . 79
4.B Principles of regression methods . . . 82
4.B.1 Linear regression with Elastic Net regularization (glmnet) . . . 82
4.B.2 Conditional Inference Tree regression (ctree2) . . . 82
4.B.3 Random Forest regression (cforest) . . . 82
4.B.4 Deep learning (keras) . . . 83

5 Bendability **85**
5.1 Introduction . . . 87
5.2 Data analysis . . . 87
5.2.1 Bendability in relation to volume fraction of phases . . . 87
5.2.2 Bendability in relation to combinations of phases . . . 90
5.2.3 Bendability in relation to number of phases . . . 90
5.3 Prediction of bendability . . . 92
5.3.1 Machine learning model performance . . . 93
5.3.2 Machine learning model interpretation . . . 94
5.4 Discussion . . . 94
5.5 Conclusions . . . 96
References . . . 97
Appendix . . . 99
5.A Raw data from literature . . . 99

III Deformation to Fracture **101**

6 Deformation properties vs. Fracture properties **103**
6.1 Introduction . . . 104
6.2 Deformation and fracture properties . . . 104
6.2.1 Tensile properties . . . 104
6.2.2 Charpy impact energy . . . 106
6.3 Relations between properties . . . 106
6.4 Prediction of Charpy impact energy . . . 108
6.5 Interpretation of random forest model . . . 110
6.5.1 Partial Dependence Plot (PDP) . . . 112
6.5.2 Accumulated Local Effects (ALE) Plot . . . 112
6.5.3 SHapley Additive exPlanations (SHAP) . . . 114
6.6 Discussion . . . 117
6.7 Conclusions . . . 118
References . . . 119
Appendix . . . 121
6.A Cleaned experimental dataset . . . 121

7 Conclusions, Discussion and Recommendations **129**
7.1 Conclusions and general discussion . . . 130
7.2 Recommendations . . . 133

Summary **137**

Samenvatting **139**

Acknowledgements **141**

Curriculum Vitæ **143**

List of Publications **145**

List of Tables

3.1 Chemical composition (wt.%) of the IF steel selected for this study. 37
3.2 Heat treatment routes with corresponding hardness values and dislocation density. Heat treatment route consists of temperature time and cooling rate, where 'Q' means quenching to room temperature, 'CR' means cold rolling. The heating rate is $10\,°C\,s^{-1}$. The cooling rate, if not stated, is $-30\,°C\,s^{-1}$. 42
3.3 Part of the detailed grain size distribution data. 43
3.A.1 Detailed descriptive statistics of the grain size for all samples. . . . 51

4.1 Reported influence of different phases on HEC for several multiphase steels. 59
4.2 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. 4.5. 65
4.3 Summary of the conditional inference tree based on the range of HEC values. 72
4.A.1 Raw data from the literature 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. 79

5.1 Results summary for linear regression between Number of phases and bendability for both separated phases (a) and with combination of all martensite and all bainite (b), corresponding to Fig. 5.5. . . . 93
5.A.1 Raw data from the literature with bendability with phase fractions in percentage and chemical content in wt%, the phases are martensite (M), ferrite (F), tempered martensite (TM), upper bainite (UB), lower bainite (LB), carbide-free bainite (CFB), bainite (B) and pearlite (P). 99

6.A.1 Cleaned dataset from public datasets [9], the area under the tensile curve is calculated from yield strength σ_y, ultimate tensile strength σ_u and uniform elongation ϵ_u by $A_d = \sigma_y * \epsilon_u + 0.5 * (\sigma_u - \sigma_y) * \epsilon_u$. 121

List of Figures

1.1 Research focus network of this thesis. The three circles represent microstructure features, deformation properties and fracture properties, respectively. The three arrows I, II and III represent the three parts of this thesis, each representing the relationship between the corresponding corners. 3

2.1 Effects of the amount of retained austenite on (a) the 0.2% yield and tensile strengths and (b) the elongation and reduction of area [10]. 10
2.2 Effect of the amount of retained austenite on the Charpy absorbed energy [10]. 10
2.3 Effects of the amount of retained austenite on fatigue limit [10]. . 11
2.4 Influence of retained austenite and carbon content on the fatigue limit (legend shows austenization temperature °C/tempering temperature °C) [11]. 12
2.5 Influence of retained austenite fraction on fatigue limit (legend shows austenization temperature °C/tempering temperature °C) (adapted from [11]). 12
2.6 Influence of σ phase on the ductility and UTS [14]. 13
2.7 Influence of σ precipitation and morphology on tensile strength with different isothermal annealing temperature [3]. 14
2.8 Influence of σ precipitation and morphology on the notch bar impact value with different isothermal annealing temperature [3]. 14
2.9 Influence of grain size on the fatigue limit [23]. 17
2.10 Hall-Petch relation from two groups of samples of AZ31 alloy with substantial difference in texture [37]. 19
2.11 Tensile elongation of the rolled AZ31 sheets as a function of pole intensity in (0002) pole figures, tensile specimens were cut at the angle of 0°, 45° and 90° to the rolling direction [39]. 20
2.12 Atomic simulation of dislocation multiplication in a Frank-Read source under external loading from the left (Adapted from [56]). 21
2.13 (a) Tensile test curve with $\sigma_{0.2}$ and yield stress determined from (b), (b) Extended Kocks-Mecking plot of the data in (a), showing a well-defined yield point [64]. 22
2.14 Influence of dislocation density and carbon content on hardness [62]. 23
2.15 Variation of the 0.2% proof stress with the square root of the dislocation density for ferritic and martensitic Fe-C alloys at 250K and 77K [59]. 24

2.16 Influence of dislocation density on yield strength [68]. 24
2.17 Relationships overview. 25

3.1 (a): Typical micro-graph of sample 'CR 3 mm'; (b): the grain boundary outline drawing of (a); (c): histogram of grain size of (a) with a log-normal fitting. 41
3.2 Grain size distribution box plot with the increasing order of hardness values from left to right. The dots represent the size of all grains measured in corresponding sample. 44
3.3 Hardness in relation to (a) mean grain size and (b) dislocation density. The dotted line in (a) is based on the linear fitting of the values for samples without cold rolling. The dotted line in (b) is based on the values of all samples. 44
3.4 LASSO plot while considering all 5 independent variables. 46
3.5 Error as a function of λ for LASSO with all 5 independent variables. The numbers on the top indicate the number of variables included in the fitted model. 46
3.A.1 Detailed descriptive statistics and corresponding histogram of the grain size for all samples. 52

4.1 Schematic illustration of HEC test after ISO 16630. The standard prescribes $D_0 = 10\,\text{mm}$. 58
4.2 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. 61
4.3 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). 62
4.4 Phase composition with increasing HEC order, (a): all the individual phase are present, (b): combining all the martensite together and all the bainite together. 63
4.5 HEC relation with number of phases (a) and with number of phases while combining all bainite and all martensite (b). 64
4.6 LASSO plot on the influence of different phase fractions on HEC. . 66
4.7 Performance comparison of all five machine learning methods. . . 69
4.8 Deep learning prediction on hole expansion capacity. 70
4.9 Variable importance plot for both conditional inference tree regression model and random forest regression model. Phases are depicted in blue, chemical elements in red. 71
4.10 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. 72

4.B.1 An example of the neural network with the input variables and output variable used in Section 4.3. 84

5.1 Schematic drawing of bending test. 88
5.2 Bendability relation with martensite (a) and ferrite (b) volume fraction. The legend represents the most abundant phase except martensite in (a) (or ferrite in (b)) for each sample. 89
5.3 Bendability relation with hard and soft phases. 90
5.4 Phase composition with increasing Bendability order, (a): all the individual phase are present, (b): combining all the martensite together and all the bainite together. 91
5.5 Bendability relation with number of phases, without combination (a) and with combination of all martensite and all bainite (b). The number shows above the median line is the average ferrite volume fraction in the corresponding group. 92
5.6 Performance comparison of all five machine learning models on the prediction of bendability. 95
5.7 Variable importance plot for both conditional inference tree regression model and random forest regression model on bendability. . . 95
5.8 The conditional inference tree regression model plot on prediction of bendability, phase fractions in percentage and chemical content in weight percent. Colors indicate the magnitude of the average bendability in the node. 96

6.1 Tensile test specimen before and after pulling (left) and tensile test process (right) (Adapted from [10]). 105
6.2 Charpy test set up and specimen dimensions (adapted from [10]). 106
6.3 Pair plot between deformation properties and fracture properties on the complete dataset. 108
6.4 Charpy impact energy colored scatter plot between UTS and UE. . 109
6.5 Relation between the estimated deformation energy and Charpy impact energy. The solid line represents the location where deformation energy equals to Charpy impact energy. 109
6.6 Random Forest predicts Charpy impact energy based on tensile test properties. 111
6.7 Error distribution of the Random Forest prediction on the testing dataset of all ten models. Two normal distributions with standard deviation of 20 J and 30 J are shown in (a). (b) shows the absolute error distribution and the accumulated distribution. 111
6.8 Accumulated Local Effects Plot of random forest model for UTS (a), YS (b), UE (c) and A_d (d) on Charpy impact energy. 115
6.9 Summary of SHAP value magnitudes over all samples for the random forest model prediction of Charpy impact energy. 117
6.10 SHAP value interactive plot between UTS and UE (a) and YS (b) for the random forest model prediction of Charpy impact energy. . . . 118

7.1 Research focus network of this thesis. The three circles represent microstructure features, deformation properties and fracture properties, respectively. The three arrows I, II and III represent the three parts of this thesis, each representing the relationship between the corresponding corners. 130

1

Introduction

Whenever an emerging structural material is developed, steel is often chosen to be the 'gold standard' for performance or quality comparison [1]. Meanwhile due to the rapidly rising requirements from various application fields, such as automotive, infrastructure and high-tech industries, steel itself as the standard is also developing fast, especially during the past decades. The magic of steel is its countless variations of microstructures and properties that come from the solid-state transformation and processing. Therefore, steel continuously gives proper combinations of microstructures and properties with the evolving requirements for safety, durability and economy [2]. Hence understanding the relationships between microstructures and properties is of great significance.

The steelmaking history can be traced back to the classical era in Ancient Iran, Ancient China, India and Rome, and exists for four millennia [3]. The modern steelmaking appears around the 17th century, when pig iron was produced from a blast furnace [4]. However, the production and processing of steel remained highly skilled art until about one century ago. The knowledge and skill of steelmaking were transferred by oral tradition for millennia, but have gradually become a matter of science over the last century.

The focus of the investigation on steels mainly falls on the effect of steelmaking and post-processing on the properties of steels for the early stage of scientific research. With the development of different characterization techniques, more and more attention is drawn to the microstructure of steels.

Within steel, automotive and related high-tech industries, there is always the persuit of high-strength steels that are better than the conventional (mild) forming steels. The higher strength leads to various advantages in weight, safety and environmental friendliness. In order to develop new steels, the steel industries make use of multi-scale microstructure modelling to predict mechanical properties from the microstructure features. Hence it is of great importance to study the contributions of microstructure features to the mechanical properties.

Microstructure is a rather complicated concept, which however can be distin-

guished by the defects. Based on the dimensions, defects in microstructures can be classified into four categories, i.e. zero-dimensional vacancies and interstitials, one-dimensional dislocations, two-dimensional grain boundaries and phase boundaries and three-dimensional phases, grains, precipitates and voids. In the numerous related research literature studies, phases, grains and dislocations are the top three mentioned microstructure features with relation to the properties of steels. A homogeneous part of a structure, which has its uniform physical and chemical characteristics, is named as phase. A constituent of a polycrystalline material, whose crystallographic orientation differs from the neighboring constituents, is called a grain. The line defect at the atomic scale in a crystal structure is termed dislocation.

Mechanical properties mainly refer to the behavior and response of materials under external loads. These are crucial for all application scenarios, and determine the range of usefulness and the services that can be expected. The deformation and fracture processes are the two main concerns of mechanical properties. Hence in this thesis, deformation properties and fracture properties are the two main focuses, for which the relations with microstructure features are discussed. The deformation of metals is achieved by the propagation of dislocations through the crystal lattice. The density of dislocations and the level of difficulty that hinders the movement of dislocations both contribute to the deformation properties. Meanwhile, fracture behavior is mainly guided by the intrinsic toughening mechanisms that are closely related to the structure and bonding mechanics of the materials. Hence fracture properties are determined not only by the kind of phases present in the material, but also by the combinations and relative volume fractions of phases. Therefore, a clear triangular structure appears just as presented in Fig. 1.1, which acts as the research focus network of this thesis.

In order to study the relation between microstructure features and mechanical properties, both of them need to be characterized. For mechanical properties, there are various testing standards for both deformation and fracture-related properties. Tensile test is the most commonly used testing method to determine deformation properties, while fracture properties are often connected with various fracture situations, such as hole expansion, bending and impact cracking. There are key parameters for each of these situations to characterize the fracture properties. For microstructure features, most of the attention is drawn to the well-known points like grain size, phase volume fraction, etc. However, there are certain aspects that are not yet generally recognized to be important, such as grain size distribution. Meanwhile, there are also microstructure features like dislocation density, which remain to be difficult to be determined precisely. Those are the complications of the study on the relations between microstructure features and mechanical properties.

This thesis aims at the development of relations between the features of multi-phase metallic microstructures of steels and the mechanical properties of the material. The quantitative characterization of the microstructure will be more involved than is now in use for estimations of the mechanical properties, which is a necessity because of the complexity of multi-phase microstructures. Moreover, the prediction of mechanical properties on the basis of microstructural features will involve

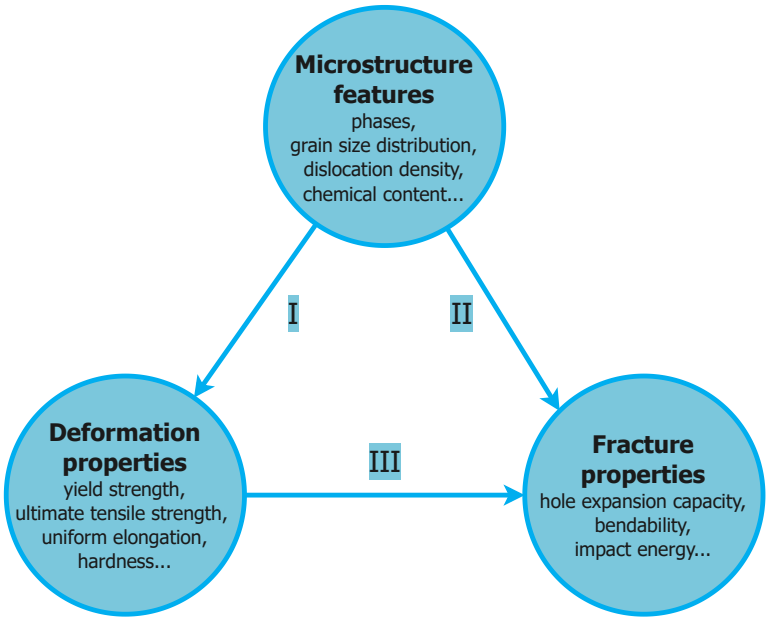


Figure 1.1: Research focus network of this thesis. The three circles represent microstructure features, deformation properties and fracture properties, respectively. The three arrows I, II and III represent the three parts of this thesis, each representing the relationship between the corresponding corners.

both deformation properties like the yield stress and fracture properties like hole expansion capacity, bendability and impact energy. Along the three arrows I, II and III in Fig. 1.1, statistical approaches are used to find relations between microstructure features and mechanical properties. Meanwhile, the possible deeply embedded relations between deformation and fracture are systematically studied.

Within this thesis, not only the experimental data obtained from the author but also materials data from the literature and online databases are used for research purposes. Meanwhile, besides the traditional statistical approaches, modern machine learning algorithms are also implemented in this thesis to accelerate the analysis and prediction of the relations. Besides the predictions, the interpretations of the machine learning algorithms are also analyzed with various techniques, which truly helps the understanding of the relationship analysis and the root behind the machine learning predictions.

The research in this thesis deepens the insight into the mechanical behaviour of the microstructure in multi-phase steels and improves property prediction, which connects the microstructure features, deformation properties and fracture properties. Results in this thesis can be directly implemented in microstructure modelling and will be directly available for researchers within the steel industry for developing new materials.

1

Outline

The research focus network of this thesis is shown in Fig. 1.1. The three anchor circles represent the three main components of this thesis, i.e. microstructure features, deformation properties and fracture properties. The three arrows indicate the parts of this thesis. Hence the thesis is divided into three parts: Microstructure to Deformation, Microstructure to Fracture and Deformation to Fracture.

Chapter 2 mainly focuses on the existing studies about various relationships between microstructure features and mechanical properties, which are summarized based on the scale of the corresponding microstructure elements, i.e. phases/precipitates, grains and dislocations. At the same time, the characteristics of these microstructure elements are also introduced.

Part I contains Chapter 3. This chapter studies the influence of microstructure features on the hardness of Interstitial Free steel. The grain size distribution and dislocation density are manipulated through different heat treatment routes combined with controlled cold rolling. With the interesting variable selection tool LASSO, the three microstructure features that most significantly influence the hardness of Interstitial Free steel are revealed.

Part II contains Chapters 4 and 5. Chapter 4 focuses on the influence of phase volume fractions and chemical content on the hole expansion capacity. Data from the literature are used in this chapter. The hole expansion capacity is analyzed in relation to individual phases, combination of phases and number of phases. Machine learning algorithms are implemented to predict the hole expansion capacity with phase volume fractions and chemical content. Moreover, the influence of microstructure features on the hole expansion capacity is also revealed.

Chapter 5 follows a similar structure as the previous chapter, but focuses on the bendability of the material. The influence of phase volume fractions on the bendability is discussed in detail. Meanwhile on the prediction of bendability from phase volume fractions and chemical content. Meanwhile, the influence of different microstructure features on bendability is revealed.

Part III contains Chapter 6. The relationships between deformation properties and fracture properties are studied. The online database offers a well-organized dataset containing the Charpy impact energy and the tensile properties. The relations between the tensile properties and the Charpy impact energy are studied in detail. Meanwhile, prediction models on the Charpy impact energy based on the tensile properties are implemented. Furthermore, with the various interpretation tools, the influence of tensile properties on the Charpy impact energy is analyzed and discussed.

The final Chapter 7 concludes the thesis and gives the recommendations for possible future work. The possible physical background and connections are also discussed from the conclusions obtained from each chapter. Future recommendations are discussed from three directions, i.e. microstructure & mechanical properties, materials data management and machine learning.

References

[1] H. Bhadeshia and R. Honeycombe, *Steels: microstructure and properties* (Butterworth-Heinemann, 2017).

[2] J. Hall and J. Fekete, *Steels for auto bodies,* in *Automotive Steels* (Elsevier, 2017) pp. 19–45.

[3] H. Akanuma, *The significance of the composition of excavated iron fragments taken from stratum iii at the site of kaman-kalehöyük, turkey,* Anatolian Archaeological Studies **14**, 147 (2005).

[4] R. F. Tylecote, *A history of metallurgy,* British Corrosion Journal **12**, 137 (1977).

2

Microstructure influence on mechanical properties

In the material sciences these are and have been,
and are most surely likely to continue to be heroic days.

Julius Robert Oppenheimer

This chapter summarizes the influence of microstructure on mechanical properties. In this chapter, various relationships between microstructure features and mechanical properties are classified based on the microstructure features. Since no crystal structure is without defects, the microstructure features can be classified by the defects. Defects in crystals are divided into four categories depending on the dimensions, which are three-dimensional phases, grains, precipitates and voids, two-dimensional grain boundaries and phase boundaries, one-dimensional dislocations and zero-dimensional vacancies and interstitials. This chapter follows the first two arrows in Fig. 1.1, explaining the studies on the relationship between microstructure features and both deformation and fracture properties. The summarized relationship matrix is shown at the end of this chapter in Fig. 2.17.

2

2.1. Introduction

In order to organize the effects of microstructures on mechanical properties, this chapter is going to discuss it in three steps, depending on the scale of the corresponding microstructure elements, i.e. phase, grain and dislocation. Phase and grain are two distinct concepts. Phase represents a homogeneous portion of a system that has uniform physical and chemical characteristics. Grain represents the constituent of a polycrystalline material where the crystallographic orientation varies from grain to grain [1]. Dislocation is used to refer to the defect on the atomic scale, which is the only one-dimensional defect. Based on these three scales, the relationships between microstructures and mechanical properties are summarized in the following sections.

2.2. Phase/Precipitation related

Phase is one of the most common microstructure features being investigated in relation to mechanical properties. The following sections first introduce the characteristics of phase and precipitation and then discuss the impact on mechanical properties.

2.2.1. Phase characteristics

- Physical and chemical properties

  Each phase has its own specific chemical and physical properties. The solute element concentration is one important factor for the chemical property. The physical property for a single phase in dual/multi- phase material is usually indicated by the nano-hardness test, which is the hardness test performed on a specific single phase area.

- Phase fraction and distribution

  Only for dual-phase or multi-phase material, it is meaningful to discuss the phase fraction and distribution. The volume fraction of phases is commonly determined by the area fraction of phases in the two-dimensional cross-section of the material [2]. The fraction and distribution information can be characterized by either optical microscopy or electron backscatter diffraction (EBSD).

- Phase boundary

  Phase boundaries exist when two or more phases meet each other. There is a discontinuous and abrupt change in physical and/or chemical properties. It is not a requirement that two phases have to differ in both physical and chemical properties [1].

Strictly speaking, precipitate is also one kind of phase, which forms from the original matrix phase. This process is called precipitation. In most metal alloys, the complexity of the precipitation process is directly linked to the amount of alloying elements [3]. The most important characteristics of precipitates are the volume

fraction and morphology. To describe the volume fraction and morphology of precipitates, different parameters are developed. The easiest and most common way to determine the volume fraction is to calculate the 2D area fraction of the precipitates. Jain *et al.* [4] calculate the volume fraction of the rod-shaped precipitates within a TEM foil with

$$f_{\mathrm{v}} = -\ln(1 - A_{\mathrm{A}})\left(\frac{DL}{DL + (D/2 + L)t}\right), \tag{2.1}$$

where D is the diameter that is approximated as a circular section. L is the rod length. t is the foil thickness and A_{A} is the projected area fraction of precipitates, which corresponds to the 2D area fraction.

Pohl *et al.* [3] used the form factor f to describe the precipitate morphology, which is given by

$$f = 4\pi\frac{A_{\mathrm{p}}}{P_{\mathrm{p}}^2}, \tag{2.2}$$

where A_{p} is the area of the precipitate and P_{p} is the perimeter of the precipitate. This form factor directly describes the roundness of the precipitates. It varies from 0, an ideally line-shaped precipitate, to 1, a perfect round precipitate.

2.2.2. Phase fraction influence

The phase fraction parameter is the most commonly mentioned factor in recent literature. There are many studies [5–9] focus on the dual-phase steel, i.e. effect of ferrite phase and martensite phase on mechanical behaviors. The general conclusions from these studies show the martensite phase has a positive effect on the strength of the materials, while at the same time increasing the brittleness. The ferrite phase tends to increase the ductility of the material, while not contributing to the strength. While looking at another phase, i.e. retained austenite, the story becomes more complicated. The influence of retained austenite fraction ($\gamma\%$) in stainless steel on the mechanical properties was investigated by Nakagawa and Miyazaki [10]. This study shows a straightforward linear relation between the amount of retained austenite and yield strength, tensile strength, elongation, reduction of area and impact energy. The linear relations are shown in Figs. 2.1 and 2.2 with the approximating equations

$$\begin{aligned}
\text{Yield strength (MPa)} &= 1192.3 - 13.6 \times f_\gamma,\\
\text{Tensile strength (MPa)} &= 1250.1 - 9.3 \times f_\gamma,\\
\text{Elongation (\%)} &= 12.16 + 0.43 \times f_\gamma,\\
\text{Reduction of area (\%)} &= 64.25 + 0.14 \times f_\gamma,\\
\text{Charpy absorbed energy (J)} &= 72.5 + 0.8 \times f_\gamma.
\end{aligned} \tag{2.3}$$

Besides these linear relations, the relationship between retained austenite fraction $f_\gamma(\%)$ and fatigue limit is shown in Fig. 2.3. The fatigue limit decreases gradually with the increase of retained austenite fraction. Different results were found by Shanmugam *et al.* [11] as shown in Fig. 2.4. Since the result in [11] shows

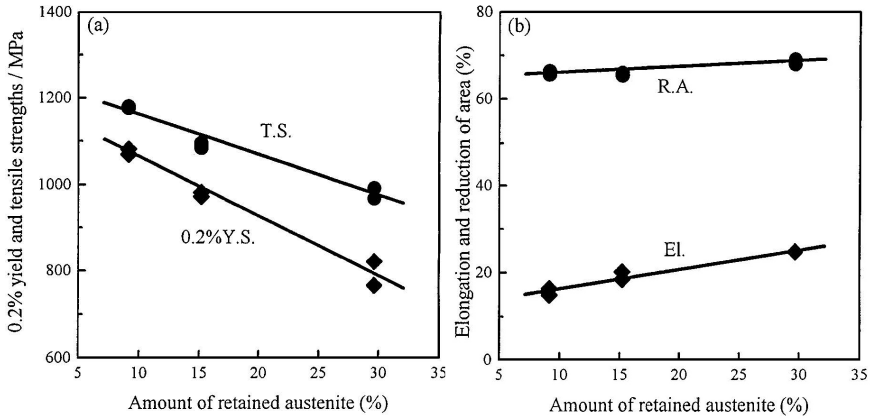


Figure 2.1: Effects of the amount of retained austenite on (a) the 0.2% yield and tensile strengths and (b) the elongation and reduction of area [10].

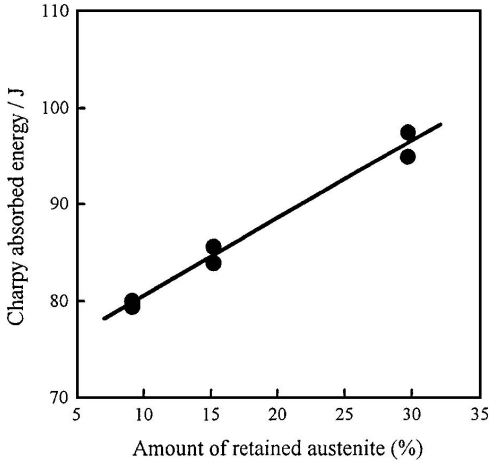


Figure 2.2: Effect of the amount of retained austenite on the Charpy absorbed energy [10].

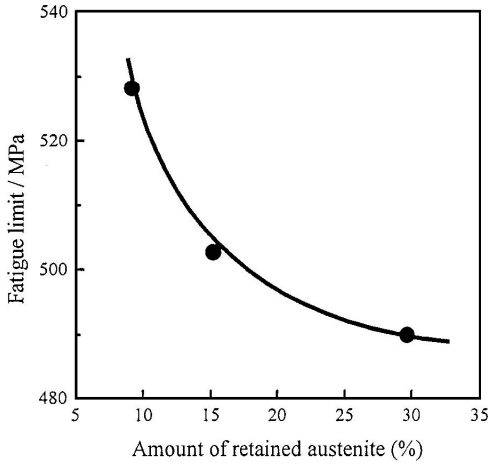


Figure 2.3: Effects of the amount of retained austenite on fatigue limit [10].

the retained austenite fraction and carbon content product, the data is extracted from [11] and the relationship between fatigue limit and retained austenite fraction is shown in Fig. 2.5. It can be seen that two contradictory results have been discovered on the relationship between retained austenite fraction and fatigue limit. Different opinions have been shown in different conditions, in ductile iron with high carbon content which hinders the strain-induced martensite formation [11], the increase of fatigue limit is due to the high-strain hardening nature of the retained austenite. The formation of persistent slip bands will be delayed, which delays the nucleation of a fatigue crack. In transformation induced plasticity steel [12], the reason is therefore the formation of very fine martensite particles in the local plastic deformation zone near the crack. Also, the finely dispersed martensite particles retard the propagation of crack. In martensitic precipitation hardening stainless steel [10], no reason was given to the decrease of fatigue limit. It may be the reason that the ductile retained austenite in the martensitic matrix plays as the crack initiation point.

There are many studies focusing on the influence of phases on certain mechanical properties. Here only a few typical examples are shown. An indication of how the phase fractions have different impact on mechanical behavior is shown. Compared to the study of the changing of mechanical properties based on changing of certain phase fractions, more focus should be on the combination of different phases, hence the investigation about the influence of multiple phases. For Advanced High-Strength Steels (AHSS), it is common to have more than two phases appearing in the microstructure. In order to better control the mechanical properties, the interaction among those present phases would make a significant role in controlling the mechanical behavior.

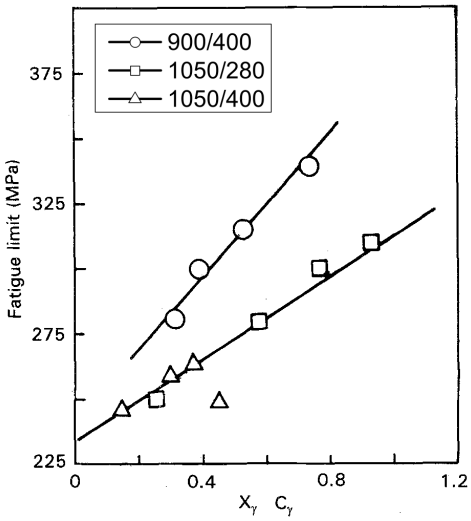


Figure 2.4: Influence of retained austenite and carbon content on the fatigue limit (legend shows austenization temperature °C/tempering temperature °C) [11].

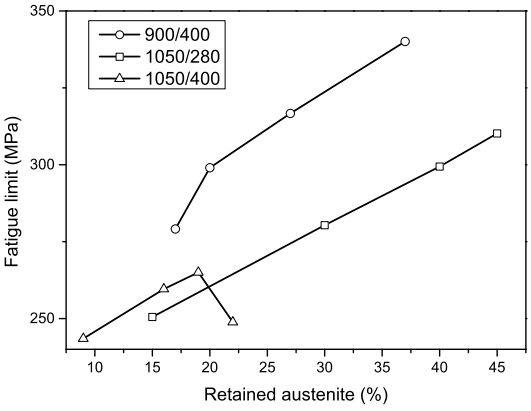


Figure 2.5: Influence of retained austenite fraction on fatigue limit (legend shows austenization temperature °C/tempering temperature °C) (adapted from [11]).

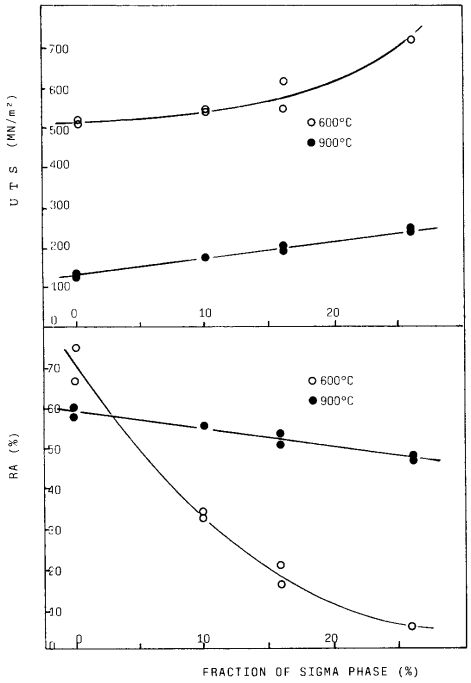


Figure 2.6: Influence of σ phase on the ductility and UTS [14].

2.2.3. Precipitation influence

In steel structures, there are various kinds of precipitates. A popular hardening mechanism is called precipitation hardening, which is the treatment to increase the strength of the materials by introducing second phase particles, i.e. precipitates. Here the focus will be on the well-known σ precipitates, which exist in various kinds of stainless steels with Cr content over 20 wt% [13]. The influence of σ precipitates on the mechanical behavior has been long time discussed. It is generally recognized that σ precipitate is a brittle phase that will decrease the toughness and elongation. The presence of the σ precipitates is the main cause of the stainless steel degradation [13].

Li *et al.* investigated the influence of σ precipitates on the ductility and UTS as shown in Fig. 2.6. Similar results on the tensile strength and notch impact toughness acquired by Pohl *et al.* are shown in Figs. 2.7 and 2.8. With the increase of volume fraction of σ precipitates, both tensile strength and UTS are increased while ductility and impact toughness are decreased. The abovementioned form factor f for 750 °C, 850 °C and 950 °C with the σ precipitate volume fraction of 15 vol% are 0.36, 0.58 and 0.61 respectively, which means the higher annealing temperature leads to more round-shape of the σ precipitates. This indicates that the rounder the shape of the σ precipitates, the less the reduction of impact toughness and the less influence on tensile strength.

The study of precipitates is often mixed with those of focus on phase influences.

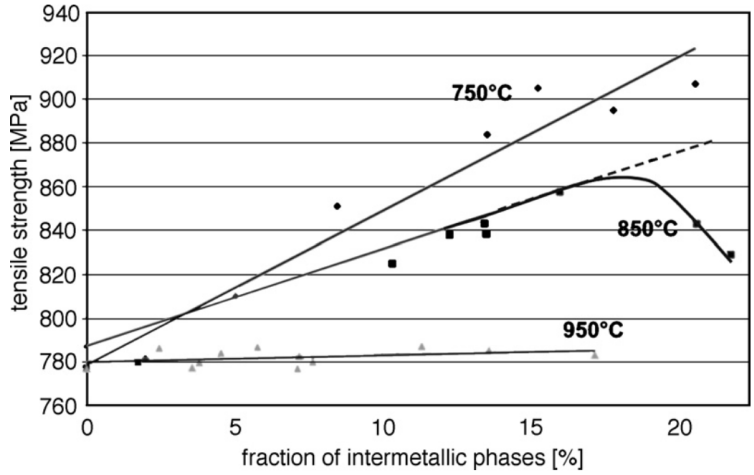


Figure 2.7: Influence of σ precipitation and morphology on tensile strength with different isothermal annealing temperature [3].

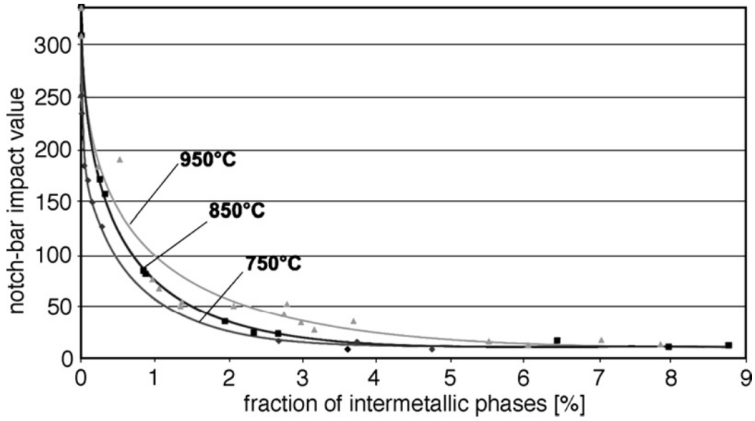


Figure 2.8: Influence of σ precipitation and morphology on the notch bar impact value with different isothermal annealing temperature [3].

It is reasonable to combine these two, but also interesting to look at precipitates at a different angle. The study of precipitates should not be limited to the size and volume, but focusing also on the shape, morphology and spatial distribution of the precipitates.

2.3. Grain related

Within the phase, grains are the next scale to investigate the materials, which has a pronounced influence on mechanical behavior. The following sections will first introduce the characteristics of grains and then discuss the impact on mechanical properties.

2.3.1. Grain characteristics

- Grain size and grain size distribution

  The most common known feature of the grain is the grain size. In order to determine grain size, there are mainly three different techniques. The first two only determine the average grain size, which are the intercept method and the grain size number method. The intercept method uses the straight lines drawing on the micrograph with the same interline distance. The average grain size is determined by the total line length divided by the number of grains that are drawing through. The grain size number method determines the grain size by comparing the micrograph under specific magnification ($100\times$) with the standard comparison charts, which are delivered by the American Society for Testing and Materials (ASTM). The third one determines the individual grain size based on the grain area, which is calculated either indirectly from metallography picture analysis or directly from Electron backscatter diffraction (EBSD). This method assumes that the grains are circular and grain size is the circular diameter.

  For grain size distribution, the size of the individual grains must be determined, hence among those three methods, only the last one could determine the grain size distribution. There is not yet a methodology that clearly shows how to describe the grain size distribution. Therefore in Chapter 3, a detailed study in grain size distribution is presented.

- Grain boundary

  Similar to phase boundary, grain boundaries form when two or more grains that show a difference in crystallographic orientation meet each other. Grain boundary is usually only several atom distances wide, in which the atomic mismatch is present due to the orientation change from one grain to the other.

  Grain boundaries could be classified into tilt or twist boundaries depending on the relationship of misorientation angle and the boundary, either vertical for a tilt boundary or parallel for a twist boundary. Depending on the degree of misorientation, tilt boundary could be further classified into low angle grain

boundary or high angle grain boundary. There is a special case called twin boundary when the two adjacent grains are in mirror lattice symmetry.

Grain boundary has its energy which is a function of the degree of misorientation. The total grain boundary energy is closely related to the grain size since the latter determines the total area of the grain boundary. The investigation on grain boundary characteristics contributing to mechanical properties is called "grain boundary engineering", which is proposed in the early 1980s [15].

- Texture

  A polycrystalline material has a "Texture" if the grains have a non-random distribution of crystallographic orientations. One simple example is the material after rolling, which usually has a rolling texture. This is due to the alignment of certain specific crystallographic planes parallel to the rolling surface and the certain specific direction in that plane parallel to the rolling direction.

  Texture could only be characterized by the use of x-rays to probe the average texture of a specimen before the early 1980s. This technique is limited to only giving the volume fraction of the grains which have a particular orientation without the information of where those grains are located in the material. The modern approach to texture was introduced in the early 1990s, which is called *microtexture*. The technique used to determine microtexture is EBSD, which is ideal to combine microstructure with crystallographic orientation information [16, 17].

2.3.2. Grain size influence

Among the characteristics of a grain, grain size is the most commonly mentioned factor in relation to mechanical properties. Almost 70 years ago, in the early 1950s, Hall and Petch proposed the well-known Hall-Petch relation [18, 19], which gives the relation of yield strength σ_y and average grain size d at ambient temperature as

$$\sigma_y = \sigma_{0y} + k_y \cdot d^{-\frac{1}{2}}, \tag{2.4}$$

where σ_{0y} is known as the friction stress for the dislocation movement within the polycrystalline grains while k_y is the Hall-Petch slope describing the local stress needed at the grain boundary for plastic flow. Eq. (2.4) also applies to cleavage fracture stress at low temperatures.

For the explanation of the Hall-Petch relation, the most widely accepted model is still the dislocation pile-up model, which is adopted in the paper of both Hall and Petch [18, 19]. This theory states that yielding occurs when the stress concentration at the grain boundary reaches the critical value. Hence the Hall-Petch relation is explained with

$$\sigma_y = M \cdot \tau_0 + M \cdot (\tau_c G b)^{\frac{1}{2}} \cdot d^{-\frac{1}{2}}, \tag{2.5}$$

where M is the Taylor orientation factor. τ_0 is friction shear stress. τ_c is the critical shear stress. G is the shear modulus and b is the length of the Burgers vector.

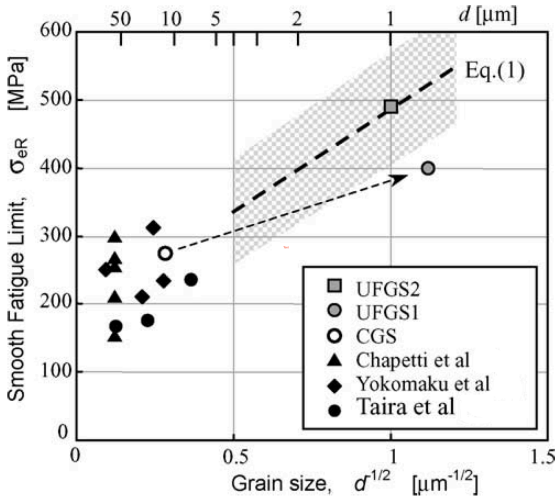


Figure 2.9: Influence of grain size on the fatigue limit [23].

The grain size influence is account for the grain boundaries, which have the effect of blocking, generating and absorbing dislocations. Hall-Petch relation explains the blocking effect of grain boundaries on gliding dislocations.

Beside the dislocation pile-up model, there are also some other models which could explain the Hall-Petch relationship. Li proposed the boundary source model considering that grain boundary ledges are the source to the dislocations, which form dislocation forest inside the grain [20]. Dislocation density model based on two types of dislocations, statistically stored dislocations (SSD) and geometrically necessary dislocations (GND) [21], assumes that SSDs are accumulated during uniform deformation while GNDs accumulate during non-uniform deformation. The work hardening model assumes that the average slip distance of glide dislocations is comparable to the grain size proposed by Meakin and Petch [22].

With the discovery of the Hall-Petch relation, a similar relation of grain size contributing to hardness was found by Hall [24], as well as the ductile-brittle transition in steel by Petch [25], the fracture toughness properties by Petch and Armstrong [26], fatigue limit and low temperature creep by Armstrong [27]. Although Eq. (2.4) was shown to be valid in many cases, there is also evidence that shows deviations from Hall-Petch relation, especially at the ultra-fine/nano grain size region [28].

Chapetti *et al.* [23] combine the data from previous researchers and their own research and give the results about grain size influence on the smooth (unnotched) fatigue limit σ_{eR} as shown in Fig. 2.9. This data collection gives the following relationship between smooth fatigue limit (MPa) and grain size (μm) in the form of Hall-Petch relation as

$$\sigma_{\mathrm{eR}} = (180 \pm 80) + \frac{300}{\sqrt{d}}. \tag{2.6}$$

2.3.3. Grain size distribution influence

In the early stage of the development of the Hall-Petch relation, the majority of researches focuses only on the mean grain size while the grain size distribution is not well recognized. Researchers either neglect the influence of the grain size distribution or assume one kind of grain size distribution function like log-normal, which is frequently mentioned [29–32]. The coefficient of variation (CV) was proposed by Kurzydłowski and Bucki [33], which involves the standard deviation (SD) of grain size, meanwhile, they also stated that grain size range is also important to mechanical properties besides the mean grain size. Berbenni *et al.* [34, 35] take the consideration of relative dispersion parameter, defined as $(D_{\max} - D_{\min})/D_{\text{mean}}$, which includes the influence of grain size range. But there are just some qualitative trends without a clear quantitative relationship discovered between the relative dispersion parameter and mechanical behavior. Lehto *et al.* proposed the volume-weighted average grain size d_{v}, which indicates that the contribution of each grain to the strength of the material is proportional to the volume of the grain. The volume-weighted average grain size is defined and adapted in the Hall-Petch relation as follows [36]:

$$d_{\text{v}} = \frac{1}{V_{\text{T}}} \sum_{i=1}^{n} V_i d_i, \tag{2.7}$$

$$d_{\text{v}}^{-1/2} = d^{-1/2}(c + f\frac{\Delta d}{d}), \tag{2.8}$$

$$\sigma_{\text{y}} = \sigma_{0\text{y}} + k_{\text{y}} d_{\text{v}}^{-1/2} = \sigma_{0\text{y}} + k_{\text{y}} d^{-1/2}(1 + f\frac{\Delta d}{d}), \tag{2.9}$$

where V_{T} is the total volume of the material and V_i is the volume of grains corresponding to the grain size d_i. Lehto *et al.* found that the volume-weighted average grain size has the relation with mean grain size as Eq. (2.8) shows, where the c and f are the constants obtained from linear regression and $c \approx 1.0$ based on the calculation. Meanwhile, the grain size dispersion parameter here is modified from Berbenni *et al.* as: $\Delta d/d = (P_{99\%} - P_{1\%})/D_{\text{mean}}$, which is aimed to minimize measurement uncertainty.

2.3.4. Texture influence

Since texture is the crystallographic orientation distribution of a polycrystalline material, which indicates the degree of anisotropy of the material, it is also one of the reasons for the anisotropy of the mechanical behavior. The influence of texture on mechanical properties is usually analyzed qualitatively, e.g. [37, 38]. Wang *et al.* [37] show the combination of grain size and texture effect on the yield stress of commercial Mg–3Al–1Zn alloy (AZ31) as shown in Fig. 2.10, where both groups follow the Hall-Petch relation nicely but have a significant difference of about 50 MPa, which is due to the texture difference induced by the different forming process.

Meanwhile, Kim *et al.* [39] give further quantitative information on texture based on the pole intensity and the relationship with tensile elongation of AZ31 as shown in Fig. 2.11. Similar result was also reported by Iwanaga *et al.* [40]. Kim *et al.*

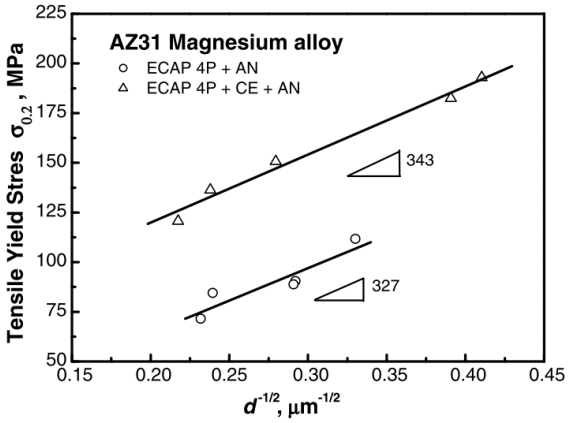


Figure 2.10: Hall-Petch relation from two groups of samples of AZ31 alloy with substantial difference in texture [37].

also recommend further detailed experimental research to find out the relation between texture intensity and ductility. Besides the increase of tensile elongation, the reduction of (0002) texture intensity also lowers the yield strength [39, 40].

The study on the grain size effect has been popular for decades. There is also the inverse Hall-Petch effect observed for nanocrystalline materials below a certain grain size by many studies [41–47]. Despite the numerous investigations about the grain size effect, not much attention was paid closely on the distribution of grain size. This was limited by the characterization techniques before. But now with the EBSD or 3D-EBSD, it is much easier to get the complete information on grain size and its distribution. There will be much more to explore with the control of grain size distribution and the corresponding mechanical behaviors. The same goes for the texture, shape and morphology of grains.

2.4. Dislocation related

Dislocation was first used to refer to a defect on the atomic scale by Taylor in 1934 [48]. The dislocation is characterized by its line element **s** and Burgers vector **b**. The line element is the unit vector in the tangential direction of the dislocation line and the Burgers vector corresponds to the length and direction of the local atomic displacement. The most important characteristics of dislocations are the density and the structure.

2.4.1. Dislocation structure

Since dislocations are defects of crystalline structure at the atomic scale, high-resolution imaging techniques are needed to reveal the dislocations. Based on difficult specimen preparation, edge and screw dislocations could be revealed under high-resolution transmission electron microscopy (HRTEM) and scanning tunneling microscopy (STM) respectively. With the conventional TEM, dislocations appear as

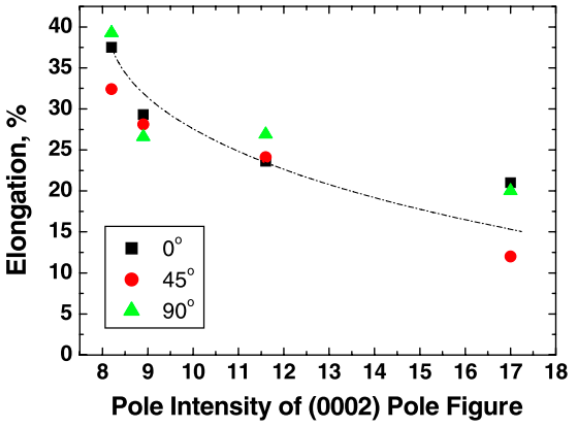


Figure 2.11: Tensile elongation of the rolled AZ31 sheets as a function of pole intensity in (0002) pole figures, tensile specimens were cut at the angle of 0°, 45° and 90° to the rolling direction [39].

dark lines in the bright field image, due to the distortion of crystal lattice around the dislocation core. Besides these techniques, dislocations can also be easily visualized by optical microscopy by the so-called etch pits, which are obtained by preferential etching on polished crystal planes [49]. This etch pits technique can be traced back to the work of Horn [50] and Gevers *et al.* [51] in 1952. The places, where dislocations intersect with the polished surface, are etched by a special etching solution.

The effective segment length is the length between the pinning points as shown in Fig. 2.12. The pinning points can be other intersected dislocations, solute elements or precipitates [52]. The dislocation effective segment length could be calculated along with the dislocation density, which will be explained in the following section.

2.4.2. Dislocation density

As the name expresses, dislocation density indicates the degree of accumulation of dislocations, which is defined by the total length of the dislocation lines per unit volume [53]. The unit for dislocation density is m^{-2}. Dislocation density can be determined by the dislocation spacing, which is determined by the intersection of dislocation line with the crystal surface while assuming all the dislocations are straight and parallel. If d is the dislocation spacing, dislocation density ρ could be calculated by [53]:

$$\rho = \frac{1}{d^2}. \tag{2.10}$$

The dislocation density in extensively annealed and in perfectly grown crystals are found to be in the order of $10^{10}\,m^{-2}$, which is caused by the crystal growth process [53]. In the heavily deformed materials, the dislocation density could reach values on the order of $10^{16}\,m^{-2}$ [54], which is normally produced by cold working

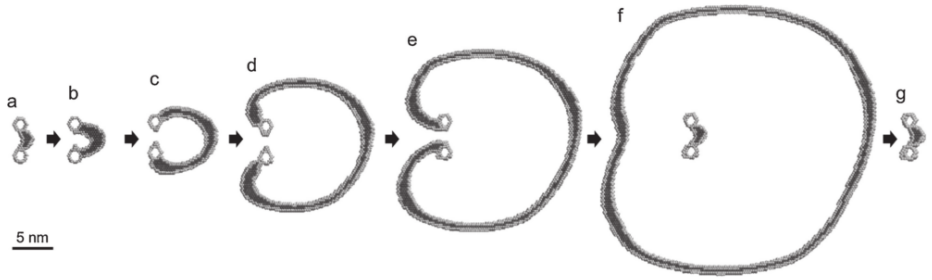


Figure 2.12: Atomic simulation of dislocation multiplication in a Frank-Read source under external loading from the left (Adapted from [56]).

[1]. The most important dislocation multiplication mechanism is the Frank-Read source [55], which is well simulated at the atomic scale by Shimokawa and Kitada [56] in Fig. 2.12. In order to retain the crystalline structure of the material and due to the internal stress field caused by dislocations, dislocation density cannot be infinitely high.

Measuring the dislocation density

The mean dislocation spacing introduced in Eq. (2.10) cannot be used in reality due to the extremely complicated distribution state of dislocations in real materials. There are several methods that can be used to measure the dislocation density.

- Measurement based on TEM:

  By using transmission electron microscopy (TEM), dislocations can be revealed under certain diffraction conditions [16, 57, 58] due to the lattice distortion around the dislocation line. Since dislocations appear as dark lines in the bright field image, the total length in the measured volume can be calculated which leads to the dislocation density. Since it is difficult to ensure that all the dislocations could be revealed in the TEM micrograph and the dislocation density may change due to the preparation of the extremely thin specimen, as well as the difficulties in determining the exact thickness of the specimen [59, 60], there are limitations to this method.

- Measurement based on XRD:

  The characterization of dislocation density using X-ray diffraction (XRD) is mainly based on the peak broadening effect induced by the strain of dislocations. The well-known accessible methods of XRD line profile analysis are the modified Williamson-Hall (MWH) method and the modified Warren-Averbach (MWA) method [61, 62]. An improved analysis method proposed by HajyAkbary *et al.* [63] combines these two methods and gives a unique value of dislocation density.

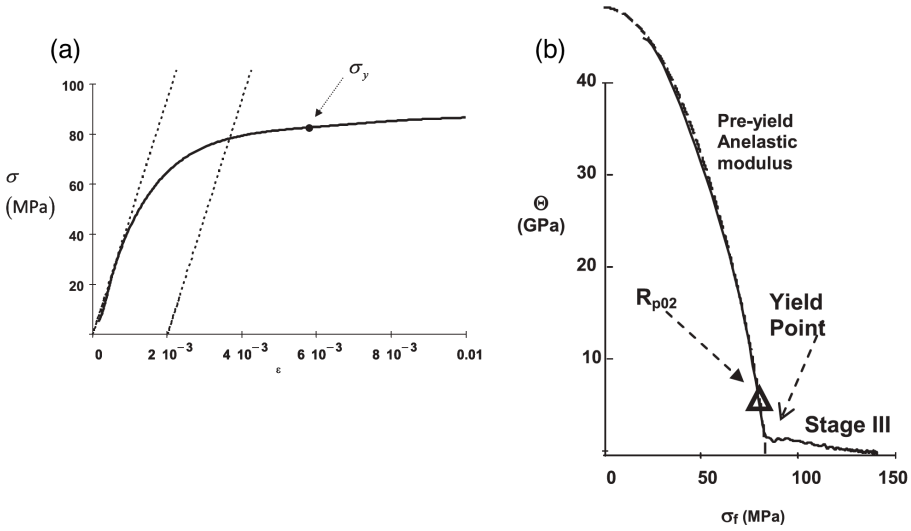


Figure 2.13: (a) Tensile test curve with $\sigma_{0.2}$ and yield stress determined from (b), (b) Extended Kocks-Mecking plot of the data in (a), showing a well-defined yield point [64].

- Measurement based on anelastic deformation behavior

  Since the anelastic deformation is caused by dislocations within the material, Arechabaleta *et al.* [52, 64] developed a method based on the extended Kocks-Mecking plot, which is shown in Fig. 2.13(b). In addition to the determination of the accurate measurement of the yield point, both dislocation density and effective segment length can be calculated.

2.4.3. Dislocation density influence

The general principle for plastic deformation is the movement of dislocations. Materials are strengthened if the dislocation motion is hindered, which could be achieved by different methods, such as grain size reduction, solid solution anchoring dislocations and strain hardening increasing the dislocation density [65–67]. However, due to the complexity in dislocation density determination, the quantitative relations between dislocation density and mechanical properties are less shown in the literature compared to other microstructure features. Cong and Murata studied the dislocation density on a low carbon steel with the relation to Vickers hardness as shown in Fig. 2.14. The Vickers hardness increases with the increase of dislocation density and carbon content. This is due to the strengthening of the martensite phase with high dislocation density as the carbon content increases [62].

Kehoe and Kelly [59] gave the relation of the 0.2% proof stress with the square root of the dislocation density, which is determined from micrographs of thin foils, for ferritic and martensitic Fe-C alloys at 250 K and 77 K as shown in Fig. 2.15, which shows a clear linear relationship. The so-called Taylor equation [48] describes the

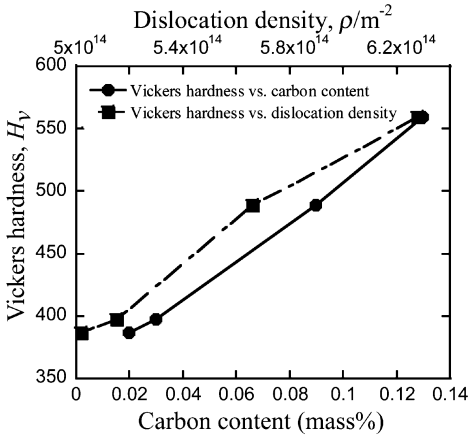


Figure 2.14: Influence of dislocation density and carbon content on hardness [62].

relation as

$$\sigma_{\mathrm{y}} = \sigma_0 + \alpha M G b \sqrt{\rho}, \tag{2.11}$$

where α is a constant, that usually ranges from 0.15 to 0.4. M is the Taylor factor. G is the shear modulus and b is the length of the Burgers vector. Arechabaleta *et al.* [68] present a cheap and easy way to determine the yield strength and dislocation density, which is also mentioned above, and the relationship is shown in Fig. 2.16, which gives an accurate experimental validation of the Taylor equation.

Nevertheless, there is still much more to investigate about the dislocation density and structure on the statistical relationship, as well as the physical background. While changing the dislocation density, it is not always possible to keep other microstructure features the same, such as grain size. Hence it is reasonable to combine the effect of dislocation influence with other microstructure features.

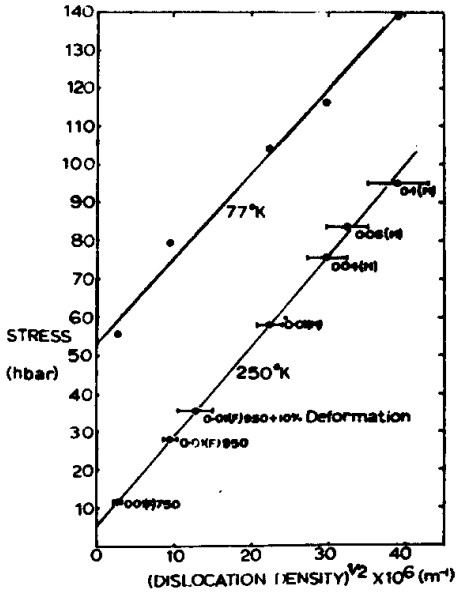


Figure 2.15: Variation of the 0.2% proof stress with the square root of the dislocation density for ferritic and martensitic Fe-C alloys at 250K and 77K [59].

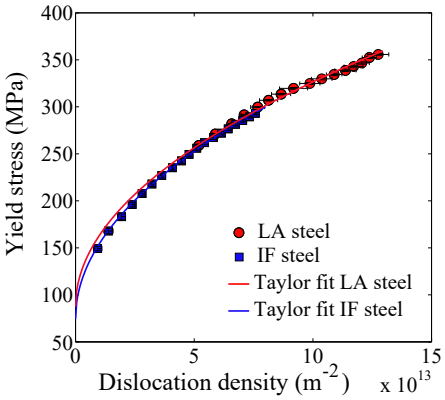


Figure 2.16: Influence of dislocation density on yield strength [68].

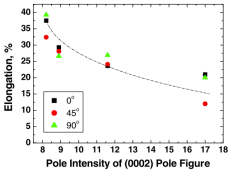
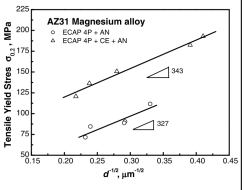
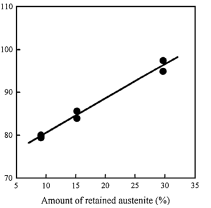
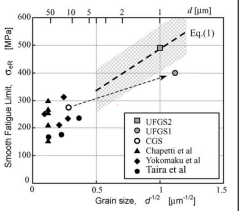
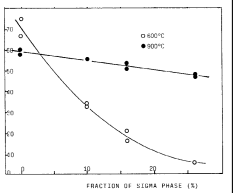
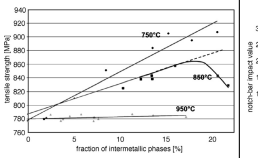
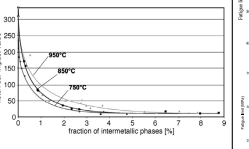
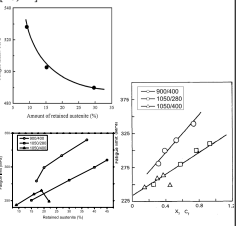
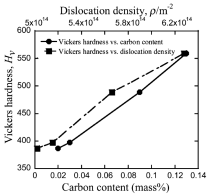
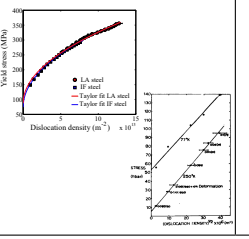
Mic-to-Mec	Ductility	Strength	Toughness	Fatigue
Grain	Tensile elongation of the rolled AZ31 sheets as a function of pole intensity in (0002) pole figures, tensile specimens were cut at the angle of 0°, 45° and 90° to the roll ing direction [39] 	Hall-Petch relation from two groups of samples of AZ31 alloy with substantial difference in texture [37] 	Charpy absorbed energy with relation to amount of retained austenite [10] 	Relationships between fatigue limit and grain size [23] 
Phase/ Precipitation	Variation of tensile properties at 600°C and 900 °C with volume fraction of σ phase [14] 	Influence of σ precipitation and morphology on tensile strength with different isothermal annealing temperature [3] 	Influence of σ precipitation and morphology on the notch bar impact value with different isothermal annealing temperature [3] 	[10, 11] 
Dislocation	Relation between dislocation density and carbon content with hardness in as-quenched low carbon 10Cr-5W steel [62] 	[59, 68] 		

Figure 2.17: Relationships overview.

References

[1] W. D. Callister, *Materials science and engineering: an introduction*, 7th ed. (John Wiley & Sons, New York, 2007).

[2] R. Haynes, *Quantitative microscopy,* in *Optical Microscopy of Materials* (Springer US, Boston, MA, 1984) pp. 88–93.

[3] M. Pohl, O. Storz, and T. Glogowski, *Effect of intermetallic precipitations on the properties of duplex stainless steel,* Materials Characterization **58**, 65 (2007).

[4] J. Jain, P. Cizek, W. Poole, and M. Barnett, *Precipitate characteristics and their effect on the prismatic-slip-dominated deformation behaviour of an Mg?6 Zn alloy,* Acta Materialia **61**, 4091 (2013).

[5] S. Tavares, P. Pedroza, J. Teodósio, and T. Gurova, *Mechanical properties of a quenched and tempered dual phase steel,* Scripta Materialia **40**, 887 (1999).

[6] A. Karelova, C. Krempaszky, E. Werner, P. Tsipouridis, T. Hebesberger, and A. Pichler, *Hole Expansion of Dual-phase and Complex-phase AHS Steels - Effect of Edge Conditions,* steel research international **80**, 71 (2009).

[7] J. H. Kim, M. Lee, D. Kim, D. Matlock, and R. Wagoner, *Hole-expansion formability of dual-phase steels using representative volume element approach with boundary-smoothing technique,* Materials Science and Engineering: A **527**, 7353 (2010).

[8] C. Bos, M. Mecozzi, and J. Sietsma, *A microstructure model for recrystallisation and phase transformation during the dual-phase steel annealing cycle,* Computational Materials Science **48**, 692 (2010).

[9] M. Calcagnotto, D. Ponge, E. Demir, and D. Raabe, *Orientation gradients and geometrically necessary dislocations in ultrafine grained dual-phase steels studied by 2D and 3D EBSD,* Materials Science and Engineering A **527**, 2738 (2010).

[10] H. Nakagawa and T. Miyazaki, *Effect of retained austenite on the microstructure and mechanical properties of martensitic precipitation hardening stainless steel,* Journal of Materials Science **34**, 3901 (1999).

[11] P. Shanmugam, P. Prasad Rao, K. Rajendra Udupa, and N. Venkataraman, *Effect of microstructure on the fatigue strength of an austempered ductile iron,* Journal of Materials Science **29**, 4933 (1994).

[12] M. Abareshi and E. Emadoddin, *Effect of retained austenite characteristics on fatigue behavior and tensile properties of transformation induced plasticity steel,* Materials & Design **32**, 5099 (2011).

[13] C.-C. Hsieh and W. Wu, *Overview of Intermetallic Sigma (σ) Phase Precipitation in Stainless Steels,* ISRN Metallurgy **2012**, 1 (2012).

[14] J. Li, T. Wu, and Y. Riquier, *δ phase precipitation and its effect on the mechanical properties of a super duplex stainless steel,* Materials Science and Engineering: A **174**, 149 (1994).

[15] T. Watanabe, *An approach to grain-boundary design for strong and ductile polycrystals,* Res Mechanica **11**, 47 (1984).

[16] O. Engler and V. Randle, *Handbook Of Texture Analysis* (CRC Press, 2009).

[17] D. J. Dingley and V. Randle, *Microtexture determination by electron back-scatter diffraction,* Journal of Materials Science **27**, 4545 (1992).

[18] E. O. Hall, *The Deformation and Ageing of Mild Steel: III Discussion of Results,* Proceedings of the Physical Society. Section B **64**, 747 (1951).

[19] N. Petch, *The cleavage strength of polycrystals,* J. Iron Steel Inst **174**, 25 (1953).

[20] J. C. Li, *Petch relation and grain boundary sources,* Transactions of the Metallurgical Society of AIME **227**, 239 (1963).

[21] M. Ashby, *The deformation of plastically non-homogeneous materials,* Philosophical Magazine **21**, 399 (1970).

[22] J. Meakin and N. Petch, *Symposium on the role of substructure in the mechanical behaviour of metals*, Tech. Rep. (ASD-TDR-63-234, Air Force Systems Command, Wright Patterson Air Force Base, Ohio, 1963).

[23] M. Chapetti, H. Miyata, T. Tagawa, T. Miyata, and M. Fujioka, *Fatigue strength of ultra-fine grained steels,* Materials Science and Engineering: A **381**, 331 (2004).

[24] E. Hall, *Variation of hardness of metals with grain size,* Nature **173**, 948 (1954).

[25] N. Petch, *The ductile-brittle transition in the fracture of α-iron: I,* Philosophical Magazine **3**, 1089 (1958).

[26] N. Petch and R. Armstrong, *Work-hardening in cleavage fracture toughness,* Acta metallurgica **37**, 2279 (1989).

[27] R. Armstrong, *The influence of polycrystal grain size on several mechanical properties of materials,* Metallurgical and Materials Transactions B **1**, 1169 (1970).

[28] A. Giga, Y. Kimoto, Y. Takigawa, and K. Higashi, *Demonstration of an inverse Hall–Petch relationship in electrodeposited nanocrystalline Ni–W alloys through tensile testing,* Scripta Materialia **55**, 143 (2006).

[29] C. Su and X. Su, *Impact of grain size and grain size distribution on the resistivity of metal nanocrystalline systems,* Computational Materials Science **108**, 62 (2015).

[30] Y. Liu, J. Zhou, and X. Ling, *Impact of grain size distribution on the multi-scale mechanical behavior of nanocrystalline materials,* Materials Science and Engineering: A **527**, 1719 (2010).

[31] B. Zhu, R. Asaro, P. Krysl, and R. Bailey, *Transition of deformation mechanisms and its connection to grain size distribution in nanocrystalline metals,* Acta Materialia **53**, 4825 (2005).

[32] T. Quested and A. Greer, *The effect of the size distribution of inoculant particles on as-cast grain size in aluminium alloys,* Acta Materialia **52**, 3859 (2004).

[33] K. Kurzydłowski and J. Bucki, *Flow stress dependence on the distribution of grain size in polycrystals,* Acta Metallurgica et Materialia **41**, 3141 (1993).

[34] S. Berbenni, V. Favier, and M. Berveiller, *Micro-macro modelling of the effects of the grain size distribution on the plastic flow stress of heterogeneous materials,* Computational Materials Science **39**, 96 (2007).

[35] N. Nicaise, S. Berbenni, F. Wagner, M. Berveiller, and X. Lemoine, *Coupled effects of grain size distributions and crystallographic textures on the plastic behaviour of IF steels,* International Journal of Plasticity **27**, 232 (2011).

[36] P. Lehto, H. Remes, T. Saukkonen, H. Hänninen, and J. Romanoff, *Influence of grain size distribution on the Hall-Petch relationship of welded structural steel,* Materials Science and Engineering: A **592**, 28 (2014).

[37] J. T. Wang, D. L. Yin, J. Q. Liu, J. Tao, Y. L. Su, and X. Zhao, *Effect of grain size on mechanical property of Mg-3Al-1Zn alloy,* Scripta Materialia **59**, 63 (2008).

[38] S. Zaefferer, J. Ohlert, and W. Bleck, *A study of microstructure, transformation mechanisms and correlation between microstructure and mechanical properties of a low alloyed TRIP steel,* Acta Materialia **52**, 2765 (2004).

[39] W. Kim, J. Lee, W. Kim, H. Jeong, and H. Jeong, *Microstructure and mechanical properties of Mg-Al-Zn alloy sheets severely deformed by asymmetrical rolling,* Scripta Materialia **56**, 309 (2007).

[40] K. Iwanaga, H. Tashiro, H. Okamoto, and K. Shimizu, *Improvement of formability from room temperature to warm temperature in AZ-31 magnesium alloy,* Journal of Materials Processing Technology **155-156**, 1313 (2004).

[41] A. Chokshi, A. Rosen, J. Karch, and H. Gleiter, *On the validity of the hall-petch relationship in nanocrystalline materials,* Scripta metallurgica **23**, 1679 (1989).

[42] U. Erb, *Electrodeposited nanocrystals: synthesis, properties and industrial applications,* Nanostructured Materials **6**, 533 (1995).

[43] G. Nieman, J. Weertman, and R. Siegel, *Micro-hardness of nanocrystalline palladium and copper produced by inert-gas condensation,* Scripta metallurgica **23**, 2013 (1989).

[44] K. Lu, W. Wei, and J. Wang, *Microhardness and fracture properties of nanocrystalline ni-p alloy,* Scripta metallurgica et materialia **24**, 2319 (1990).

[45] G. Nieman, J. Weertman, and R. Siegel, *Mechanical behavior of nanocrystalline cu and pd,* Journal of Materials Research **6**, 1012 (1991).

[46] P. G. Sanders, J. Eastman, and J. Weertman, *Elastic and tensile behavior of nanocrystalline copper and palladium,* Acta materialia **45**, 4019 (1997).

[47] A. Giga, Y. Kimoto, Y. Takigawa, and K. Higashi, *Demonstration of an inverse Hall-Petch relationship in electrodeposited nanocrystalline Ni-W alloys through tensile testing,* Scripta Materialia **55**, 143 (2006).

[48] G. I. Taylor, *The Mechanism of Plastic Deformation of Crystals. Part I. Theoretical,* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **145**, 362 (1934).

[49] G. F. Vander Voort, *Metallography, principles and practice* (ASM International, 1999).

[50] F. H. Horn, *CXX. Screw dislocations, etch figures, and holes,* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **43**, 1210 (1952).

[51] R. Gevers, S. Amelinckx, and W. Dekeyser, *Interfero metric study of etchpits,* Die Naturwissenschaften **39**, 448 (1952).

[52] Z. Arechabaleta, P. van Liempt, and J. Sietsma, *Quantification of dislocation structures from anelastic deformation behaviour,* Acta Materialia **115**, 314 (2016).

[53] G. Gottstein, *Physical Foundations of Materials Science* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).

[54] Y. B. Wang, J. C. Ho, Y. Cao, X. Z. Liao, H. Q. Li, Y. H. Zhao, E. J. Lavernia, S. P. Ringer, and Y. T. Zhu, *Dislocation density evolution during high pressure torsion of a nanocrystalline Ni–Fe alloy,* Applied Physics Letters **94**, 091911 (2009).

[55] F. C. Frank and W. T. Read, *Multiplication Processes for Slow Moving Dislocations,* Physical Review **79**, 722 (1950).

[56] T. Shimokawa and S. Kitada, *Dislocation Multiplication from the Frank-Read Source in Atomic Models,* Materials Transactions **55**, 58 (2014).

[57] B. Fultz and J. M. Howe, *Transmission Electron Microscopy and Diffractometry of Materials* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 1–758, arXiv:arXiv:1011.1669v3 .

[58] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy* (Springer US, Boston, MA, 2009).

[59] M. Kehoe and P. Kelly, *The role of carbon in the strength of ferrous martensite,* Scripta Metallurgica **4**, 473 (1970).

[60] S. Morito, J. Nishikawa, and T. Maki, *Dislocation Density within Lath Martensite in Fe-C and Fe-Ni Alloys,* ISIJ International **43**, 1475 (2003).

[61] T. Ungár and A. Borbély, *The effect of dislocation contrast on x-ray line broadening: A new approach to line profile analysis,* Applied Physics Letters **69**, 3173 (1996).

[62] Z. Cong and Y. Murata, *Dislocation Density of Lath Martensite in 10Cr-5W Heat-Resistant Steels,* Materials Transactions **52**, 2151 (2011).

[63] F. HajyAkbary, J. Sietsma, A. J. Böttger, and M. J. Santofimia, *An improved X-ray diffraction analysis method to characterize dislocation density in lath martensitic structures,* Materials Science and Engineering: A **639**, 208 (2015).

[64] P. van Liempt and J. Sietsma, *A physically based yield criterion I. Determination of the yield stress based on analysis of pre-yield dislocation behaviour,* Materials Science and Engineering: A **662**, 80 (2016).

[65] G. Zlateva and Z. Martinova, *Microstructure of Metals and Alloys* (CRC Press, 2008).

[66] H. Bhadeshia and R. Honeycombe, *Steels: microstructure and properties* (Butterworth-Heinemann, 2017).

[67] D. Raabe, B. Sun, A. Kwiatkowski Da Silva, B. Gault, H.-W. Yen, K. Sedighiani, P. Thoudden Sukumar, I. R. Souza Filho, S. Katnagallu, E. Jägle, P. Kürnsteiner, N. Kusampudi, L. Stephenson, M. Herbig, C. H. Liebscher, H. Springer, S. Zaefferer, V. Shah, S.-L. Wong, C. Baron, M. Diehl, F. Roters, and D. Ponge, *Current Challenges and Opportunities in Microstructure-Related Properties of Advanced High-Strength Steels,* Metallurgical and Materials Transactions A **51**, 5517 (2020).

[68] Z. Arechabaleta, P. van Liempt, and J. Sietsma, *Unravelling dislocation networks in metals,* Materials Science and Engineering: A **710**, 329 (2018).

I

Microstructure to Deformation

3

The combined influence of grain size distribution and dislocation density on hardness of interstitial free steel

It can scarcely be denied that the supreme goal of all theory
is to make the irreducible basic elements
as simple and as few as possible
without having to surrender the adequate representation
of a single datum of experience.

Albert Einstein

Understanding the relationship between microstructure features and mechanical properties is of great significance for the improvement and specific adjustment of steel properties. The relationship between mean grain size and yield strength is established by the well-known Hall-Petch equation. But due to the complexity of the grain configuration within materials, considering only the mean value is unlikely to give a complete representation of the mechanical behavior. The classical Taylor equation is often used

This chapter is based on the article: **W. Li**, M. Vittorietti, G. Jongbloed, J. Sietsma, *The combined influence of grain size distribution and dislocation density on hardness of interstitial free steel*, Journal of Materials Science & Technology **45**, 35 (2019).

to account for the effect of dislocation density, but not thoroughly tested in combination with grain size influence. In the present study, systematic heat treatment routes and cold rolling followed by annealing are designed for Interstitial Free (IF) steel to achieve ferritic microstructures that not only vary in mean grain size, but also in grain size distribution and in dislocation density, a combination that is rarely studied in the literature. Optical microscopy is applied to determine the grain size distribution. The dislocation density is determined through XRD measurements. The hardness is analyzed on its relation with the mean grain size, as well as with the grain size distribution and the dislocation density. With the help of the variable selection tool LASSO, it is shown that dislocation density, mean grain size and kurtosis of grain size distribution are the three features that most strongly affect the hardness of IF steel.

keywords

Interstitial free steel, Hardness, Grain size distribution, Dislocation density, LASSO

3.1. Introduction

The most commonly known feature of the grain configuration in metallic microstructures is the grain size, which is also the most commonly mentioned factor in relation to mechanical properties among all characteristics of grains. Almost 70 years ago, in the early 1950s, Hall and Petch [1, 2] proposed the by now well-known Hall-Petch relation, which gives the relation of yield strength σ_{y} and average grain size μ_{d} at ambient temperature as:

$$\sigma_{\mathrm{y}} = \sigma_{0\mathrm{y}} + k_{\mathrm{y}} \cdot {\mu_{\mathrm{d}}}^{-1/2}, \tag{3.1}$$

where $\sigma_{0\mathrm{y}}$ is known as the friction stress for the dislocation movement within the grains in a polycrystalline microstructure, while k_{y} is the Hall-Petch slope describing the local stress needed at the grain boundary for plastic flow [3]. Since the early stage of the development of the Hall-Petch relation, the majority of studies focused only on the mean grain size while the effect of the grain size distribution is not well recognized. Researchers either neglect the additional information that may be present in the grain size distribution, or assume one kind of grain size distribution function like log-normal, which is frequently mentioned [4–7]. The coefficient of variation $C_V = s/\mu_{\mathrm{d}}$ of the grain size distribution was proposed by Kurzydłowski and Bucki [8] to represent the grain size distribution, which involves the standard deviation (s) and mean (μ_{d}) grain size. They state that, beside the mean grain size, the grain size range is important to model mechanical properties. Berbenni *et al.* [9], Nicaise *et al.* [10] consider a relative dispersion parameter, defined as $p_{\mathrm{dis}} = (d_{\max} - d_{\min})/\mu_{\mathrm{d}}$, which includes the influence of the grain size range, i.e. largest grain size $d_{\max}$ and smallest grain size $d_{\min}$. But just some qualitative trends are proposed between the relative dispersion parameter and mechanical behavior without a clear quantitative relationship. Lehto *et al.* [11] propose to use the volume-weighted average grain size d_{v}, which indicates that the influence of each grain on the strength of the material is proportional to the volume of the grain. The volume-weighted average grain size d_{v} is defined and included in a Hall-Petch-like relation as follows:

$$d_{\mathrm{v}} = \frac{1}{V_{\mathrm{T}}} \sum_{i=1}^{n} V_i d_i, \tag{3.2}$$

$$d_{\mathrm{v}}^{-1/2} = \mu_{\mathrm{d}}^{-1/2} (c + f \cdot p_{\mathrm{dis_m}}), \tag{3.3}$$

$$\sigma_{\mathrm{y}} = \sigma_{0\mathrm{y}} + k_{\mathrm{y}} \mu_{\mathrm{d}}^{-1/2} (c + f \cdot p_{\mathrm{dis_m}}), \tag{3.4}$$

where V_{T} is the total volume of the material, n is the number of grains and V_i is the volume of grains of the grain size d_i. Lehto *et al.* found that the volume-weighted average grain size has a relation with mean grain size as Eq. (3.3) shows, where c and f are constants obtained from linear regression of experimental data, which gives $c \approx 1.0$. The grain size dispersion parameter here is modified from Berbenni *et al.* [9] as: $p_{\mathrm{dis_m}} = (P_{99\%} - P_{1\%})/\mu_{\mathrm{d}}$. $P_{99\%}$ and $P_{1\%}$ are the grain sizes at 99% and 1% probability levels, which is more robust than $d_{\max}$ and $d_{\min}$. The modified Hall-Petch equation is proposed as Eq. (3.4) by combining Eq. (3.1) and

Eq. (3.3). Similar to the volume-weighted average grain size, Raeisinia and Sinclair [12] propose a representative grain size D_{R}, with relation to mean grain size as follows:

$$D_{\mathrm{R}} = \mu_{\mathrm{d}} \exp\left(\frac{9S^2}{4}\right), \tag{3.5}$$

$$\sigma_{\mathrm{y}} = \sigma_{0\mathrm{y}} + k_{\mathrm{y}} \exp\left(\frac{-9S^2}{8}\right) \mu_{\mathrm{d}}^{-1/2}, \tag{3.6}$$

where S is the standard deviation of the assumed log-normal grain size distribution function. Eq. (3.6) is the modified Hall-Petch equation with the representative grain size incorporated. It helps in explaining why the Hall-Petch slope changes in some cases. Here the representative grain size is calculated based on the assumption that the grain size follows a log-normal distribution function.

General principle for plastic deformation is the movement of dislocations. Materials are strengthened if the dislocation motion is hindered, which can be achieved by different methods, such as grain size reduction, precipitation, solid solution anchoring dislocations and strain hardening by increasing the dislocation density. However, due to the complexity of accurate dislocation density determination, the quantitative relations between dislocation density and mechanical properties are less validated in the literature compared to other microstructure features, such as grain size. Taylor [13] proposed a relation between the 0.2% proof stress σ_{y} and the square root of the dislocation density $\sqrt{\rho}$ as follows:

$$\sigma_{\mathrm{y}} = \sigma_0 + \alpha M G b \sqrt{\rho}, \tag{3.7}$$

where α is a constant, which usually ranges from 0.15 to 0.4, M is the Taylor factor, G is the shear modulus and b is the length of the Burgers vector. Cong and Murata studied the dislocation density in low carbon steel in relation to Vickers hardness, which shows that the Vickers hardness increases with the increase of dislocation density and carbon content. This is due to the strengthening of martensite phase with high dislocation density as the carbon content increases and the solid solution strengthening effect of carbon [14]. Kehoe and Kelly [15] determine the dislocation density ρ from TEM micro-graphs of thin foils, for ferritic and martensitic Fe-C alloys at 250 K and 77 K, which shows a clear linear relationship between σ_{y} and $\sqrt{\rho}$. Arechabaleta *et al.* [16] present an accessible and easy way to determine the yield strength and dislocation density, which gives an accurate experimental validation of the Taylor equation and physical interpretation of the parameter α.

It is shown above that many researchers have shown that grain size distribution does have pronounced influence on the strength of materials. But all the mentioned studies concerning grain size distribution have their drawbacks. None of them actually indicates which parameters are the most important to describe the strength effect of the grain size distribution. In the present paper, in order to establish a relationship between grain size distribution, dislocation density and mechanical properties, interstitial free (IF) steel is chosen to be the steel type to manipulate the grain size distribution and dislocation density, minimizing the influence from

Table 3.1: Chemical composition (wt.%) of the IF steel selected for this study.

Fe	C	Ti	Mn	Al	Si	Cr	Si	P
Bal.	0.005	0.081	0.077	0.055	0.052	0.016	0.011	0.003

other aspects, such as carbon content, precipitates and different phases. The ferrite grain size distribution and dislocation density are controlled through different deformation and heat treatment routes. The study results in a combined relation of mean grain size, grain size distribution and dislocation density with hardness.

3.2. Material and Methods

3.2.1. Material and processing

The chemical composition of the interstitial free steel is shown in Table 3.1. Different heat treatment routes were applied on the as received hot-rolled IF steel plate with sample dimensions of 10 mm × 4 mm × 2.5 mm. The accurate temperature control during the heat treatment was achieved by a Bähr dilatometer model DIL 805A/D. The detailed heat treatment routes can be seen from Table 3.2, where the heat treatment temperature, time and cooling rate are indicated. 'Q' stands for quenching to room temperature, as measured in the dilatometer, during which the cooling rate above 300 °C is around $-200\,°C\,s^{-1}$. Heating and cooling rate during the heat treatment, if not stated in the table, were controlled at $10\,°C\,s^{-1}$ and $-30\,°C\,s^{-1}$, respectively.

In order to introduce the influence of dislocations, cold rolling was performed on the 4 mm thick hot-rolled plate. The plate was cold rolled to the final thickness of 3.00 mm, 1.70 mm and 1.13 mm, causing a strain of 25%, 58% and 72%, respectively. The 1.13 mm cold rolled plate was subjected to annealing at 400 °C, 450 °C and 600 °C for time ranging from 1.5 min to 75 min, followed by quenching to room temperature with helium in the dilatometer.

3.2.2. Grain size distribution determination

The heat treated samples were mechanically ground with P800, P1200 and P2000 grit abrasive papers, followed by polishing with 3 µm and 1 µm diamond paste. In order to reveal the grain boundaries of IF steel, Marshall's Reagent was used first, with the etching time around 3 to 4 seconds, followed by a 20 second etch in 2% nital [17]. The nital etching after using Marshall's Reagent is necessary to reveal the ferrite grain boundaries. The microstructure was captured by the light optical microscope Olympus BX60M. The grain size analysis was conducted with the ImageJ software, which detects the grain boundaries based on the image contrast of the transformed binary pictures and gives the individual information of each grain, including the visible grain area A_i. The grain size value for each grain i, i.e. equivalent grain diameter d_i, was calculated by assuming that the area equals that of a perfect circle with diameter d_i, where $d_i = 2\sqrt{A_i/\pi}$. The mean grain size μ_d, standard deviation s, skewness $\tilde{\mu}_3$ and kurtosis $\tilde{\mu}_4$ of the grain size distribution

based on n grains are calculated with

$$\mu_{\mathrm{d}} = \frac{1}{n}\sum_{i=1}^{n} d_i, \tag{3.8}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \mu_{\mathrm{d}})^2}{n-1}}, \tag{3.9}$$

$$\tilde{\mu}_3 = \frac{\sum_{i=1}^{n}(d_i - \mu_{\mathrm{d}})^3}{n * s^3}, \tag{3.10}$$

$$\tilde{\mu}_4 = \frac{\sum_{i=1}^{n}(d_i - \mu_{\mathrm{d}})^4}{n * s^4} - 3. \tag{3.11}$$

These are the key values, i.e. quantitative measures, which describe the grain size distribution. Skewness represents the asymmetry of the grain size distribution, which equals zero when the distribution is perfectly symmetric. Skewness is negative when the low-value tail of the distribution is longer than the high-value tail and positive when the high-value tail of the distribution is longer than the low-value tail [18]. Kurtosis in this paper actually refers to "excess kurtosis", which is defined as kurtosis minus 3. Kurtosis measures the "tailedness" of the distribution, which is equal to zero for normal distributions, regardless of the values of its parameters. It is positive for so-called leptokurtic distribution, with fatter tails and negative for so-called platykurtic distribution, with thinner tails, compared to the normal distributions [19].

3.2.3. Dislocation density determination by X-Ray diffraction (XRD)

X-Ray diffraction measurements were conducted on a Bruker D8 Advance diffractometer with Bragg-Brentano geometry with Lynxeye position sensitive detector, operating at 45 kV, 40 mA, using Cu Kα radiation (wavelength $\lambda = 0.154\,06$ nm) without the scatter screen. The scanning speed was controlled at $0.005\,°\,s^{-1}$ between 38° and 90° and $0.013\,°\,s^{-1}$ between 90° and 152°. The reflections obtained for single phase body-centered cubic (bcc) ferrite structure are {110}, {200}, {211}, {220}, {310} and {222}. The samples were sticked with small amount of plasticine, due to the small dimension, on a Si{510} wafer holder L40SiB. The obtained data from two different scanning speed stages is merged in Bruker software Diffrac.EVA 4.2.2 and processed with X'Pert Highscore 2.2c. The correction for the instrumental broadening is done by subtraction of the measured peak width of reference sample Lanthanum Hexaboride (LaB_6)-SRM 660a. Peak fitting is processed with Topas with the split Pearson VII function, in order to obtain the peak width. The dislocation density calculation is based on the modified Williamson-Hall method, which is explained in detail in [20, 21].

3.2.4. Hardness measurement

In this study, mechanical properties are represented by the hardness value, which was measured with the Struers DualScan 70 auto hardness machine using Vickers hardness standards with 30 N force in order to cover enough grains over the sample surface to reveal the macroscopic hardness of the test sample. There were 8 individual hardness measurements conducted on each sample to minimize the test uncertainty.

3.2.5. Variable selection method: LASSO

Due to the large number of independent variables measured from microstructure features, i.e. mean, standard deviation, skewness and kurtosis of the grain size and dislocation density, and the limited data we can obtain from experiments, we employ a statistical method called Least Absolute Shrinkage and Selection Operator (LASSO), which gives the order of importance for the variables based on the influence of each microstructure variable on the target mechanical property. LASSO was first introduced in the geophysics literature by Santosa and Symes [22] and then independently rediscovered by Tibshirani [23]. LASSO, as a tool widely used in the field of machine learning, is rarely adopted in materials science studies, especially in the experimental field, to perform variable selection. The few studies using LASSO in materials science are computational studies [24–27]. Classically, if there are only a few (or only one) explanatory variables, one can use the method of least squares to estimate the regression parameters. If the number of parameters comes close to or exceeds the number of data points, the least squares estimator becomes unstable or is even not well defined anymore. LASSO is a popular method to circumvent these problems, as it adapts the least squares criterion, leading to a well-defined estimator and at the same time provides a way to select the most relevant explanatory variables from the whole set.

In this study, the $glmnet$ package in R [28] was used. The LASSO prediction of the target mechanical property f at the point $\mathbf{x}$ is

$$\hat{f}_{LASSO}(\mathbf{x}) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_j, \tag{3.12}$$

where x_j is the jth variable in the prediction point $\mathbf{x}$, which is a vector containing p variables measured on the microstructure. In order to use this technique and referring to Eqs. (3.1) and (3.7), ${\mu_\mathrm{d}}^{-1/2}$ and $\rho^{1/2}$ were used as variables. The estimate $\hat{\beta}_j$ is the corresponding coefficient in the LASSO which minimizes the objective function:

$$L(\beta_0, \beta_1, ..., \beta_p) = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda H_\mathrm{r}^2 \sum_{j=1}^{p} |\frac{\beta_j}{\beta_{j,\lambda=0}}|. \tag{3.13}$$

Here n is the number of the experimental data, x_{ij} is the jth variable of the ith microstructure, y_i is the target mechanical property measured on the sample corresponding to the microstructure with measurement $\mathbf{x}_i$, $\lambda H_\mathrm{r}^2 \sum_{j=1}^{p} |\frac{\beta_j}{\beta_{j,\lambda=0}}|$ is called

the shrinkage penalty which is controlled by the tuning parameter λ [23, 25] and in which $H_r = 1\,\text{HV}$. This tuning parameter λ determines the decrease of regression coefficients. It can be seen in Eq. (3.13) that for a positive value of λ, the fit will yield lower values for β_j than for $\lambda = 0$. This deteriorates the goodness of the fit. At very large values of λ, all fit parameters β_j will become zero. With gradually increasing λ, the fit parameters that are least important for the trends in the target parameter will decrease to zero more rapidly than the more important parameters. We will present the LASSO analysis as a plot of $\beta_j/(\beta_{j,max} - \beta_{j,min})$ vs λ that shows the change in β values with λ, with $\beta_{j,max}$ and $\beta_{j,min}$ the maximum and minimum values in the applied λ range. This will be presented and analyzed for the present study in Section 3.3.4.

3.3. Results and discussion

3.3.1. Microstructure and hardness

Based on the microstructure analysis, the 16 samples contain only the ferritic phase without any other phases, as a typical micrograph shows in Fig. 3.1(a). The XRD measurements also confirm the single phase conclusion by containing just the reflections from bcc structure. The heat treatment routes applied on the hot rolled IF steel plates, as well as the cold rolled plates, result in pronounced differences in hardness values, as shown in Table 3.2. The hardness varies from $(58.6 \pm 1.5)\,\text{HV}$ up to $(167.0 \pm 1.9)\,\text{HV}$. As shown clearly by the data, hardnesses are significantly lower for the heat treated samples without cold rolling being applied compared with those after cold rolling, which is due to higher dislocation density induced by cold rolling. The only slowly cooled sample '1000 °C 10 min 1 °C s^{-1}' has the lowest hardness of $(58.6 \pm 1.5)\,\text{HV}$ due to the largest mean grain size obtained by slow cooling and low dislocation density. Hardness increases with the increase of rolling thickness reduction and decreases with increasing annealing temperature and time, which is closely related to the dislocation density, as shown in following sections.

3.3.2. Grain size distribution

Following the microstructure image processing and analysis, the dimensions of around 200 grains are obtained from each sample. An example of grain boundary outlines and corresponding grain size distribution for sample 'CR 3 mm' is shown in Fig. 3.1(b) and (c). The detailed grain size data as well as the histograms of all samples can be seen in Table 3.A.1 and Fig. 3.A.1 in the appendix. Both grain size distribution data and histograms show that the distribution of grain size varies from sample to sample, whereas it cannot always be well fitted with a log-normal distribution function, as proposed and applied by many researchers [5, 7, 9, 29, 30]. The skewness of all samples is positive, which corresponds to a longer high-value tail of the grain size distributions. Only the sample with slow cooling '1000 °C 10 min 1 °C s^{-1}' has a negative kurtosis, which corresponds to its thinnest tails as the first histogram shows in Fig. 3.A.1 in the appendix. All other samples have positive kurtosis, which means they have fatter tails compared to the normal distribution. The box plot for grain size data of each sample can be seen in Fig. 3.2, whereas

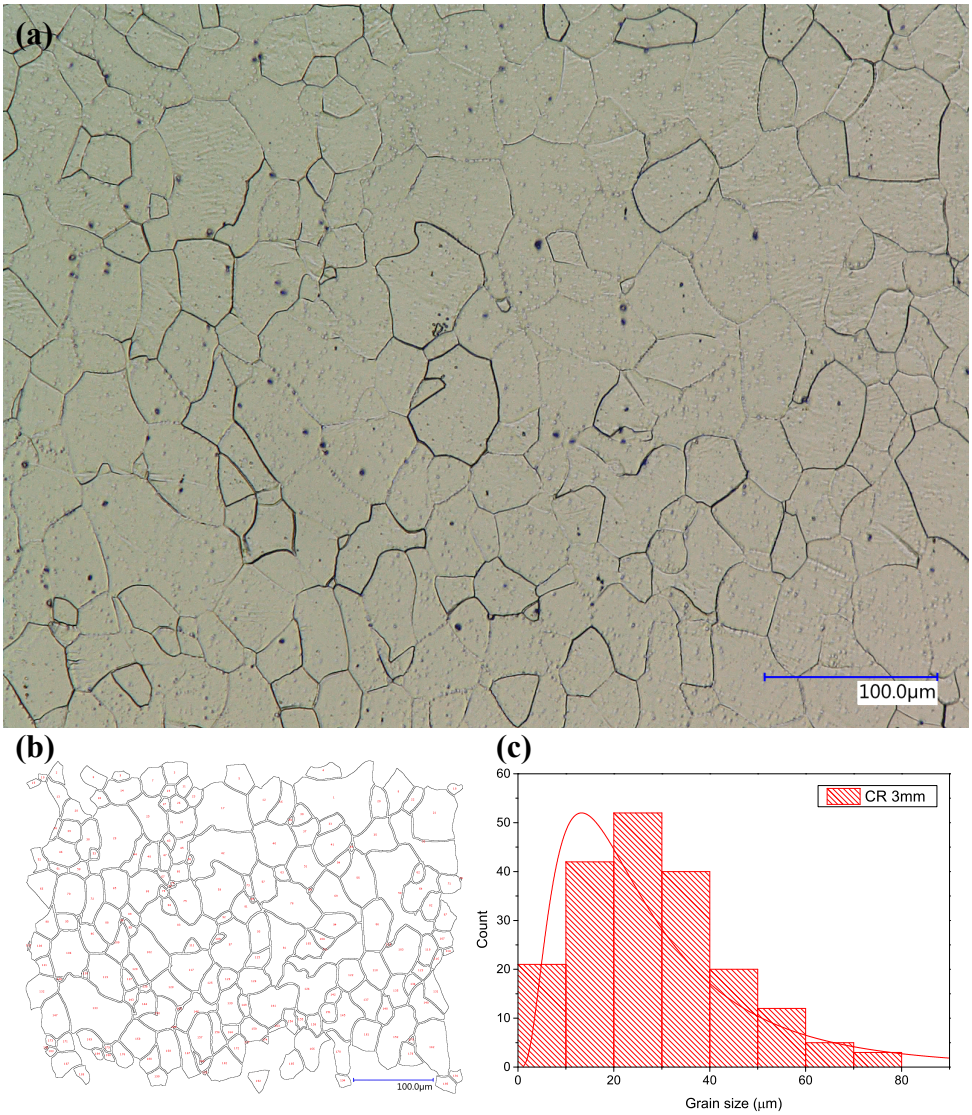


Figure 3.1: (a): Typical micro-graph of sample 'CR 3 mm'; (b): the grain boundary outline drawing of (a); (c): histogram of grain size of (a) with a log-normal fitting.

Table 3.2: Heat treatment routes with corresponding hardness values and dislocation density. Heat treatment route consists of temperature time and cooling rate, where 'Q' means quenching to room temperature, 'CR' means cold rolling. The heating rate is 10 °C s^{-1}. The cooling rate, if not stated, is −30 °C s^{-1}.

Heat treatment routes	Mean Hardness (HV)	Hardness Std. (HV)	Dislocation density (10^{12}/m^2)	Dislocation density error (10^{12}/m^2)
1000 °C 10 min 1 °C s^{-1}	58.6	1.5	1	-
1000 °C 10 min 750 °C 5 min Q	60.9	2.4	1	-
1000 °C 10 min 800 °C 5 min Q	62.0	1.3	1	-
1000 °C 10 min 850 °C 5 min Q	62.3	1.6	1	-
1000 °C 10 min 700 °C 5 min Q	64.9	0.9	1	-
1000 °C 10 min 80 °C s^{-1}	66.7	2.3	1	-
800 °C 5 min Q	68.1	0.5	1	-
1000 °C 10 min 200 °C s^{-1}	69.1	1.7	3.1	0.9
700 °C 5 min Q	75.9	3.2	1	-
400 °C 5 min Q	80.1	1.3	1	-
CR 3 mm	127.3	2.3	43	3
CR 1.7 mm	139.9	2.2	75	4
CR 1.13 mm 600 °C 75 min Q	141.8	1.6	12	2
CR 1.13 mm 450 °C 9 min Q	159.6	2.4	60	4
CR 1.13 mm 400 °C 9 min Q	163.4	2.0	75	5
CR 1.13 mm	167.0	1.9	72	4

the relation with hardness will be explained in the following section. The boxplot clearly shows that cold rolled samples and those without austenitization stage, i.e. heated up to 800 °C or lower, have lower mean grain size and narrower grain size distribution. Samples that underwent austenitization at 1000 °C have larger mean grain size and broader grain size distribution. This is also shown by the detailed data in Table 3.A.1 in the appendix.

3.3.3. Dislocation density

The dislocation density, as shown in Table 3.2, is assumed to be 1×10^{12} m^{-2} [31] for the undeformed plates which is related to the detection limit of the applied XRD method. It increases to 7.5×10^{13} m^{-2} for cold rolled plates. With the increase of strain, dislocation density increases. By elevating the temperature or extending the time of the annealing treatment after cold rolling, dislocation density generally decreases, except for the condition '1.13 mm 400 °C 9 min'.

3.3.4. Correlation of hardness with grain size distribution and dislocation density

Table 3.3: Part of the detailed grain size distribution data.

Heat treatment routes	μ_d (µm)	s (µm)	Skewness	Kurtosis
1000 °C 10 min 1 °C s^{-1}	159.1	94.0	0.68	−0.20
1000 °C 10 min 750 °C 5 min Q	135.8	95.5	1.02	0.51
1000 °C 10 min 800 °C 5 min Q	145.0	113.3	1.13	0.41
1000 °C 10 min 850 °C 5 min Q	92.5	59.0	0.88	0.20
1000 °C 10 min 700 °C 5 min Q	123.9	111.7	1.59	2.24
1000 °C 10 min 80 °C s^{-1}	71.6	53.7	0.99	0.32
800 °C 5 min Q	32.9	18.9	1.04	0.85
1000 °C 10 min 200 °C s^{-1}	122.5	70.0	0.75	0.20
700 °C 5 min Q	32.4	20.3	0.86	0.39
400 °C 5 min Q	31.0	18.8	0.81	0.45
CR 3 mm	28.3	16.0	0.63	0.22
CR 1.7 mm	24.2	20.1	1.25	1.42
CR 1.13 mm 600 °C 75 min Q	24.8	25.5	2.01	5.54
CR 1.13 mm	26.8	24.5	1.51	2.19

General trend

The correlation of grain size and hardness can be seen in Figs. 3.2 and 3.3. The boxplot in Fig. 3.2 indicates the grain size range from 25% to 75%, while the whisker line shows the 1% to 99% range. The box middle line shows the median value of grain size, while the square and the extended middle horizontal line show the mean value. The trend of increasing hardness with decreasing mean grain size is seen in Fig. 3.2, but can be found in a more quantitative manner in Fig. 3.3, where the dotted line indicates the Hall-Petch trend for the microstructures that were not plastically deformed. The relation of dislocation density and hardness is shown in Fig. 3.3, which shows the generally positive effect of dislocation density on hardness, which is linear with $\sqrt{\rho}$, as expressed in the Taylor equation (Eq. (3.7)). By combining the boxplot in Fig. 3.2 and dislocation density with hardness plot in Fig. 3.3, it is clearly shown that cold rolled samples have much higher hardness than those without cold rolling, which is due to higher dislocation density and lower mean grain size.

Overall fitting with mean grain size and dislocation density

For a comprehensive relation between microstructure and mechanical properties, the effect of mean grain size and dislocation density should be considered in combination. Least squares linear regression leads to the following fit with the adjusted R-squared of 0.91 and Root Mean Square Error (RMSE) of 11 HV:

$$H_{\mathrm{V}} = H_0 + k_{\mathrm{d}} * {\mu_{\mathrm{d}}}^{-1/2} + k_\rho * \rho^{1/2}, \tag{3.14}$$

where $k_{\mathrm{d}} = 187.4\,\mathrm{HV\mu m^{1/2}}$, $k_\rho = 1.1 \times 10^{-5}\,\mathrm{HVm}$ and $H_0 = 34.0\,\mathrm{HV}$. The weight for hardness applied in the linear regression is the inverse of standard deviation of

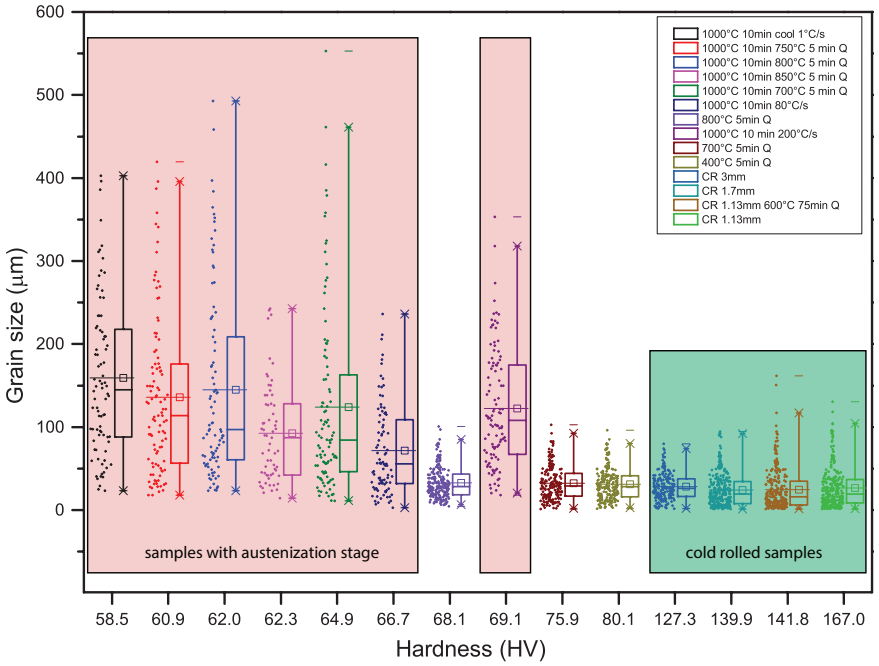


Figure 3.2: Grain size distribution box plot with the increasing order of hardness values from left to right. The dots represent the size of all grains measured in corresponding sample.

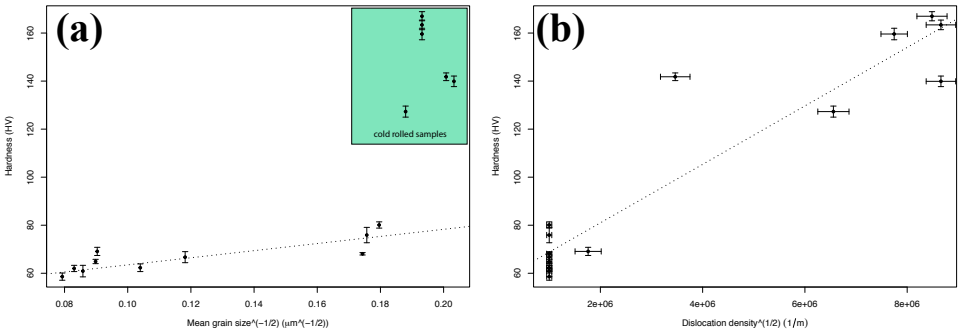


Figure 3.3: Hardness in relation to (a) mean grain size and (b) dislocation density. The dotted line in (a) is based on the linear fitting of the values for samples without cold rolling. The dotted line in (b) is based on the values of all samples.

hardness. The RMSE is calculated based on the predicted hardness H_p, individual hardness values H_i and number of samples N with:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(H_\mathrm{p} - H_\mathrm{i})^2}{N}}. \tag{3.15}$$

Variable selection by LASSO

In order to establish a clear relation among hardness, grain size distribution and dislocation density and also be able to predict hardness under similar conditions, different variables describing both grain size distribution and dislocation density are included in a linear model and the model is fitted using the LASSO method (Section 3.2.5). These variables are, as defined before, dislocation density, mean grain size, standard deviation, skewness and kurtosis of the grain size distribution, which are introduced in the LASSO Eq. (3.12). For the LASSO test, μ_d and ρ are brought into the calculation as ${\mu_\mathrm{d}}^{-1/2}$ and $\rho^{1/2}$ in order to apply the linear fitting scheme of the LASSO test equation, equivalent to Eq. (3.13). The resulting LASSO plot is shown in Fig. 3.4. In the LASSO plot, from the left to the right direction, the tuning parameter λ becomes larger, hence generates a higher penalty for the fit parameter (β) values, therefore the fit parameter values decrease, eventually to zero. The variable that maintains a non-zero coefficient until the highest value of λ is the dislocation density, which therefore is recognized as the variable that has the most significant impact on hardness. The following variables are mean and kurtosis of the grain size distribution. Skewness does not appear in this LASSO plot, which may be related to the relatively high correlation between skewness and kurtosis.

Final fitting with variables selected by LASSO

Based on the LASSO method, the mean square sum of residuals as a function of λ is shown in Fig. 3.5 with the number of non-zero fit parameters (β) along the top axis. In general, the aim of model descriptions is to balance accuracy and simplicity. The dashed line on the left in Fig. 3.5 represents the most accurate model, while the one on the right represents the simplest model with an error within the standard error of the most accurate model, i.e. the model with minimum number of coefficients which gives a good accuracy. In this case, the entire range between the most accurate model and the simplest model implies the use of three variables, which are dislocation density, mean and kurtosis.

Therefore, hardness is fitted with a combination of the Hall-Petch equation and the Taylor equation, while including the influence of the kurtosis $\tilde{\mu}_4$. The model can be expressed by:

$$H_\mathrm{V} = H_0 + k_\mathrm{d} * {\mu_\mathrm{d}}^{-1/2} + k_\rho * \rho^{1/2} + k_\mathrm{k} * \tilde{\mu}_4, \tag{3.16}$$

with adjusted R-Squared of 0.96 and RMSE of 7 HV, where $k_\mathrm{d} = 127.3\,\mathrm{HV\mu m^{1/2}}$, $k_\rho = 9.9 \times 10^{-6}\,\mathrm{HVm}$, $k_\mathrm{k} = 7.2\,\mathrm{HV}$ and $H_0 = 35.7\,\mathrm{HV}$. The positive value of k_k shows the positive effect on hardness from kurtosis of grain size distribution, which means that fatter tails of grain size distribution contribute to higher hardness. A

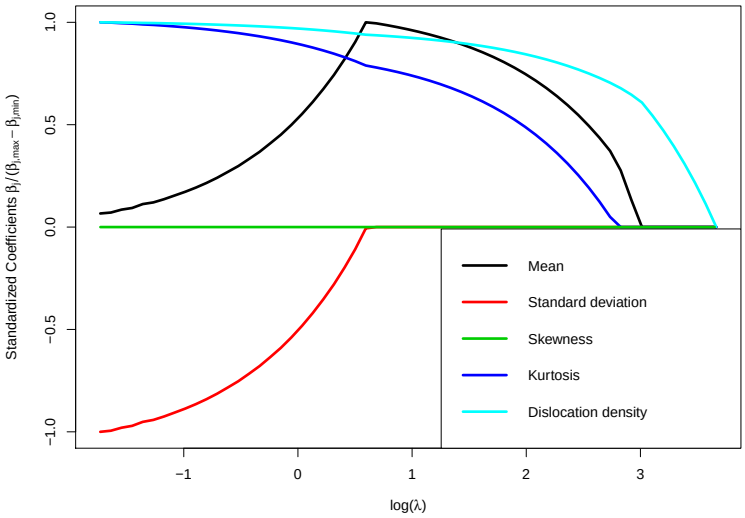


Figure 3.4: LASSO plot while considering all 5 independent variables.

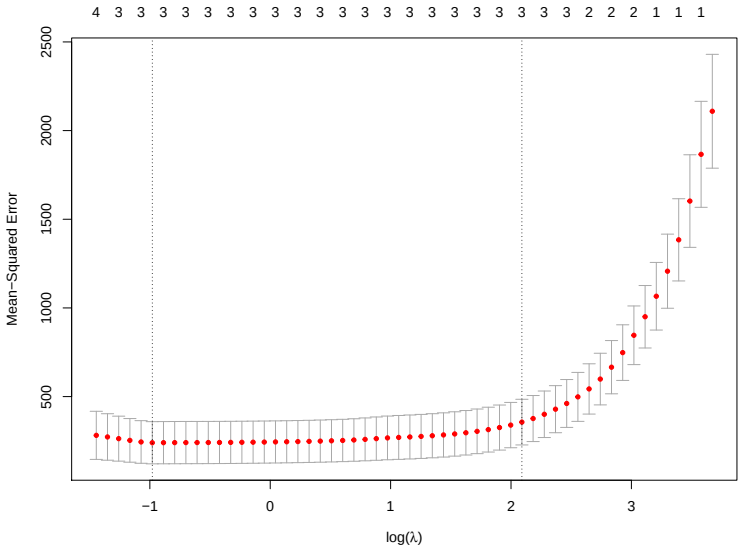


Figure 3.5: Error as a function of λ for LASSO with all 5 independent variables. The numbers on the top indicate the number of variables included in the fitted model.

microstructure which has more grains at either low or high grain size values tends to have higher hardness than a microstructure that has more grains having similar grain size values.

As increasing the number of explanatory variable generally increases R-squared and adjusted R-squared cannot represent the predictive capability of our model, in order to judge the predictive capability of our model, predicted R-squared is calculated, which is driven by the leave-one-out cross-validation [32]. Each data point in turn is removed from the dataset for the fitting and the model is refitted using the remaining data points. Then the hardness value of the removed data point is calculated using the new model, hence leads to the predicted residual error sum of squares, which calculates the predicted R-squared. When comparing the models from Eqs. (3.14) and (3.16), the predicted R-squared is increasing from 0.79 to 0.88, which means that including extra explanatory variable, i.e. kurtosis, does improve the predictive capability of our model, to an uncertainty of ± 7 HV.

Analysis of the Hall-Petch slope

From Eqs. (3.14) and (3.16), two Hall-Petch slopes have been generated: $k_{\mathrm{d1}} = 187.4\,\mathrm{HV\mu m^{1/2}}$ and $k_{\mathrm{d2}} = 127.3\,\mathrm{HV\mu m^{1/2}}$. The literature[33] indicates that the Hall-Petch slopes for pure iron and low carbon steels range between $150\,\mathrm{MPa\mu m^{1/2}}$ and $600\,\mathrm{MPa\mu m^{1/2}}$. With 50 ppm solute carbon, the Hall-Petch slope is found to be $560\,\mathrm{MPa\mu m^{1/2}}$[33], which just match the IF steel with 50 ppm carbon content in this study. Since yield strength can be determined with good precision from Vickers hardness by $\sigma_{\mathrm{y}} = H_{\mathrm{V}}/3$ [34, 35] and 1 HV = 9.8 MPa, the Hall-Petch slope of $560\,\mathrm{MPa\mu m^{1/2}}$ corresponds to $171.4\,\mathrm{HV\mu m^{1/2}}$. Therefore, the obtained value k_{d1} and k_{d2} in this study are both reasonable and comparable with the literature data.

The slope obtained while only using mean grain size and dislocation density (k_{d1}) is obviously higher than that using mean grain size, dislocation density and kurtosis (k_{d2}). This is because the Eq. (3.16) takes the grain size distribution factor kurtosis into the fitting. Since ${\mu_{\mathrm{d}}}^{-1/2}$ and $\tilde{\mu}_4$ are positively correlated (see Table 3.3) and kurtosis has a positive influence on hardness, it is a mathematical consequence that the Hall-Petch slope will be lower. This also indicates that the two aspects of the grain size distribution effect on hardness have been separated by adding the grain size distribution term kurtosis.

3.4. Conclusions

To understand how the combination of grain size distribution and dislocation density influences the hardness of IF steel, a series of IF steel plates were given different microstructures through different heat treatment routes in combination with cold rolling. Based on the microstructure characterization and hardness measurement, the following conclusions are drawn from this research.

- Different heat treatment routes and degrees of cold rolling change the dislocation density and grain size distribution, which contribute to the variation of hardness. Cold rolling plays a more significant role in increasing hardness, due to the decrease of mean grain size and increase of dislocation density.

- LASSO, as a relatively new method in experimental materials science, plays an important role as the variable selection tool, which gives further insight into the relative influence of different variables and selects for the simplest model with good accuracy.
- The combined contribution of dislocation density and grain size distribution on hardness of IF steel plates can be expressed by the equation $H_V = H_0 + k_d * \mu_d{}^{-1/2} + k_\rho * \rho^{1/2} + k_k * \tilde{\mu}_4$, where $k_d = 127.3\,\mathrm{HV\mu m^{1/2}}$, $k_\rho = 9.9 \times 10^{-6}\,\mathrm{HVm}$, $k_k = 7.2\,\mathrm{HV}$ and $H_0 = 35.7\,\mathrm{HV}$.

3

References

[1] E. O. Hall, *The Deformation and Ageing of Mild Steel: III Discussion of Results,* Proceedings of the Physical Society. Section B **64**, 747 (1951).

[2] N. Petch, *The cleavage strength of polycrystals,* J. Iron Steel Inst **174**, 25 (1953).

[3] R. W. Armstrong, *60 Years of Hall-Petch: Past to Present Nano-Scale Connections,* Materials Transactions **55**, 2 (2014).

[4] C. Su and X. Su, *Impact of grain size and grain size distribution on the resistivity of metal nanocrystalline systems,* Computational Materials Science **108**, 62 (2015).

[5] Y. Liu, J. Zhou, and X. Ling, *Impact of grain size distribution on the multi-scale mechanical behavior of nanocrystalline materials,* Materials Science and Engineering: A **527**, 1719 (2010).

[6] B. Zhu, R. Asaro, P. Krysl, and R. Bailey, *Transition of deformation mechanisms and its connection to grain size distribution in nanocrystalline metals,* Acta Materialia **53**, 4825 (2005).

[7] T. Quested and A. Greer, *The effect of the size distribution of inoculant particles on as-cast grain size in aluminium alloys,* Acta Materialia **52**, 3859 (2004).

[8] K. Kurzydłowski and J. Bucki, *Flow stress dependence on the distribution of grain size in polycrystals,* Acta Metallurgica et Materialia **41**, 3141 (1993).

[9] S. Berbenni, V. Favier, and M. Berveiller, *Micro-macro modelling of the effects of the grain size distribution on the plastic flow stress of heterogeneous materials,* Computational Materials Science **39**, 96 (2007).

[10] N. Nicaise, S. Berbenni, F. Wagner, M. Berveiller, and X. Lemoine, *Coupled effects of grain size distributions and crystallographic textures on the plastic behaviour of IF steels,* International Journal of Plasticity **27**, 232 (2011).

[11] P. Lehto, H. Remes, T. Saukkonen, H. Hänninen, and J. Romanoff, *Influence of grain size distribution on the Hall-Petch relationship of welded structural steel,* Materials Science and Engineering: A **592**, 28 (2014).

[12] B. Raeisinia and C. Sinclair, *A representative grain size for the mechanical response of polycrystals,* Materials Science and Engineering: A **525**, 78 (2009).

[13] G. I. Taylor, *The Mechanism of Plastic Deformation of Crystals. Part I. Theoretical,* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **145**, 362 (1934).

[14] Z. Cong and Y. Murata, *Dislocation Density of Lath Martensite in 10Cr-5W Heat-Resistant Steels,* Materials Transactions **52**, 2151 (2011).

[15] M. Kehoe and P. Kelly, *The role of carbon in the strength of ferrous martensite,* Scripta Metallurgica **4**, 473 (1970).

[16] Z. Arechabaleta, P. van Liempt, and J. Sietsma, *Unravelling dislocation networks in metals,* Materials Science and Engineering: A **710**, 329 (2018).

[17] B. L. Bramfitt and A. O. Benscoter, *Metallurgia Italiana*, Vol. 95 (ASM International, 2001).

[18] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures Fifth Edition* (Chapman & Hall CRC, 2011).

[19] P. H. Westfall, *Kurtosis as Peakedness, 1905–2014. R.I.P.* The American Statistician , 191 (2014).

[20] F. HajyAkbary, J. Sietsma, A. J. Böttger, and M. J. Santofimia, *An improved X-ray diffraction analysis method to characterize dislocation density in lath martensitic structures,* Materials Science and Engineering: A **639**, 208 (2015).

[21] Z. Arechabaleta, P. van Liempt, and J. Sietsma, *Quantification of dislocation structures from anelastic deformation behaviour,* Acta Materialia **115**, 314 (2016).

[22] F. Santosa and W. W. Symes, *Linear Inversion of Band-Limited Reflection Seismograms,* SIAM Journal on Scientific and Statistical Computing **7**, 1307 (1986).

[23] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso,* Journal of the Royal Statistical Society. Series B (Methodological) **58**, 267 (1996).

[24] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, *Machine learning in materials informatics: recent applications and prospects,* npj Computational Materials **3**, 54 (2017).

[25] Y. Zhang and C. Ling, *A strategy to apply machine learning to small datasets in materials science,* npj Computational Materials **4**, 25 (2018).

[26] L. M. Ghiringhelli, J. Vybiral, E. Ahmetcik, R. Ouyang, S. V. Levchenko, C. Draxl, and M. Scheffler, *Learning physical descriptors for materials science by compressed sensing,* New Journal of Physics **19**, 023017 (2017).

[27] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, *Big Data of Materials Science: Critical Role of the Descriptor,* Physical Review Letters **114**, 105503 (2015).

[28] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2013).

[29] S. Ramtani, H. Bui, and G. Dirras, *A revisited generalized self-consistent polycrystal model following an incremental small strain formulation and including grain-size distribution effect,* International Journal of Engineering Science **47**, 537 (2009).

[30] W. Woo, T. Ungár, Z. Feng, E. Kenik, and B. Clausen, *X-Ray and Neutron Diffraction Measurements of Dislocation Density and Subgrain Size in a Friction-Stir-Welded Aluminum Alloy,* Metallurgical and Materials Transactions A **41**, 1210 (2010).

[31] G. Gottstein, *Physical Foundations of Materials Science* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York, New York, NY, 2009).

[33] S. Takaki, *Review on the Hall-Petch Relation in Ferritic Steel,* Materials Science Forum **654-656**, 11 (2010).

[34] J. R. Cahoon, W. H. Broughton, and A. R. Kutzak, *The determination of yield strength from hardness measurements,* Metallurgical Transactions **2**, 1979 (1971).

[35] P. Zhang, S. Li, and Z. Zhang, *General relationship between strength and hardness,* Materials Science and Engineering: A **529**, 62 (2011).

Appendix

3.A. Descriptive statistics and histogram of all samples

The detailed grain size distribution data and all the histograms are shown here in appendix.

Table 3.A.1: Detailed descriptive statistics of the grain size for all samples.

Heat treatment	N total	Mean	Standard Deviation	SE of mean	Skewness	Kurtosis	Minimum	Median	Maximum	Range (Maximum - Minimum)	P1	P99
1000 °C 10 min 1 °C s^{-1}	84	159.1	94.0	10.3	0.68	−0.20	23.5	145.1	402.8	379.3	23.5	402.8
1000 °C 10 min 750 °C 5 min Q	105	135.8	95.5	9.3	1.02	0.51	17.9	113.9	419.6	401.6	18.0	395.6
1000 °C 10 min 800 °C 5 min Q	93	145.0	113.3	11.7	1.13	0.41	23.3	96.8	493.0	469.7	23.3	493.0
1000 °C 10 min 850 °C 5 min Q	63	92.5	59.0	7.4	0.88	0.20	14.3	87.3	242.6	228.4	14.3	242.6
1000 °C 10 min 700 °C 5 min Q	102	123.9	111.7	11.1	1.59	2.24	11.0	84.3	552.9	541.9	11.4	461.3
1000 °C 10 min 80 °C s^{-1}	93	71.6	53.7	5.6	0.99	0.32	3.0	55.7	236.1	233.1	3.0	236.1
800 °C 5 min Q	214	32.9	18.9	1.3	1.04	0.85	4.5	28.4	100.6	96.1	6.5	85.0
1000 °C 10 min 200 °C s^{-1}	107	122.5	70.0	6.8	0.75	0.20	18.2	108.3	353.2	335.0	20.6	318.1
700 °C 5 min Q	198	32.4	20.3	1.4	0.86	0.39	1.4	29.2	102.9	101.5	2.2	92.6
400 °C 5 min Q	214	31.0	18.8	1.3	0.81	0.45	1.5	28.3	96.3	94.7	3.1	80.2
CR 3 mm	195	28.3	16.0	1.1	0.63	0.22	1.6	26.4	79.8	78.2	2.1	74.3
CR 1.7 mm	273	24.2	20.1	1.2	1.25	1.42	1.3	19.4	94.5	93.2	1.5	92.2
CR 1.13 mm 600 °C 75 min Q	257	24.8	25.5	1.6	2.01	5.54	1.3	15.9	161.7	160.4	1.6	117.0
CR 1.13 mm	284	26.8	24.5	1.5	1.51	2.19	1.3	19.4	130.5	129.2	1.7	104.2

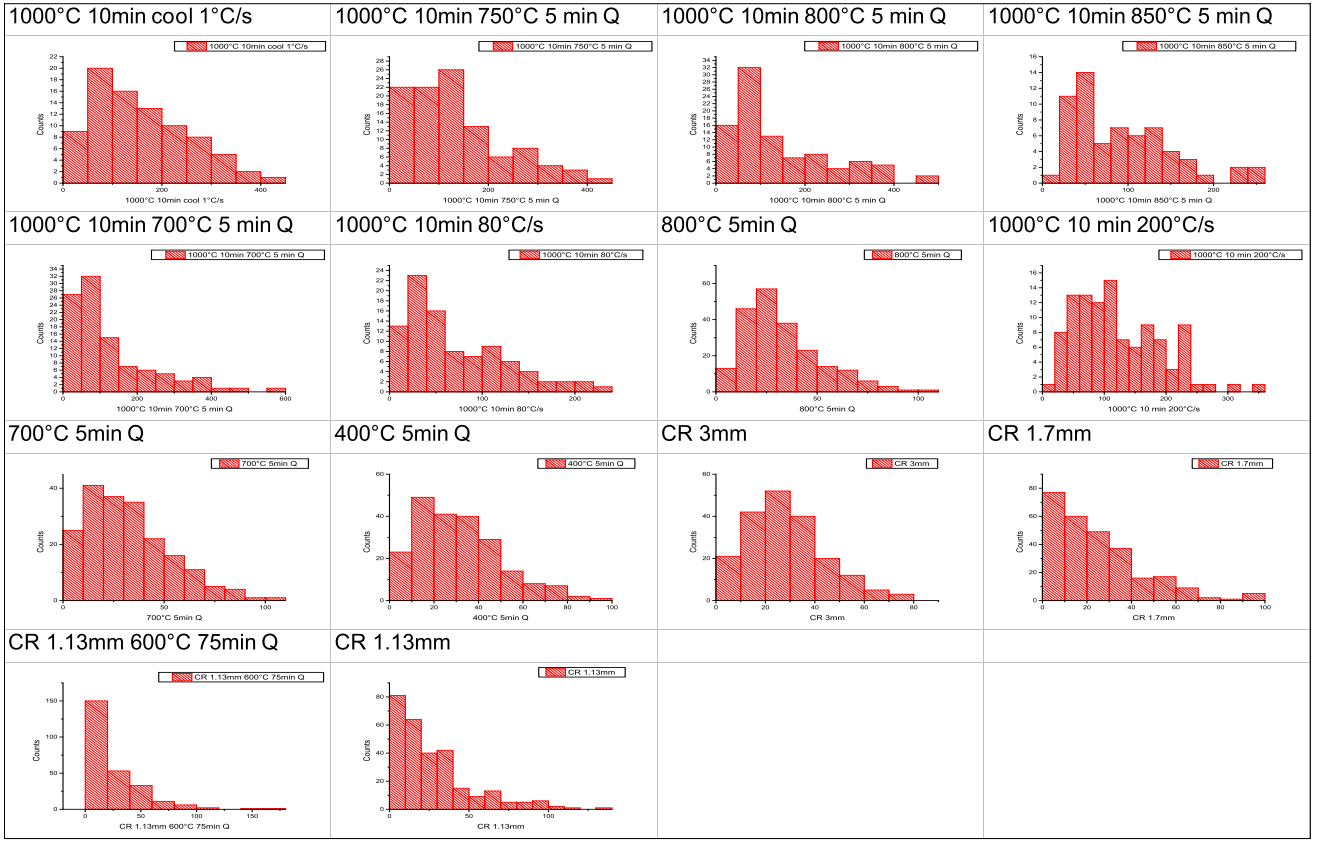


Figure 3.A.1: Detailed descriptive statistics and corresponding histogram of the grain size for all samples.

II

Microstructure to Fracture

4

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

Success in creating AI would be the biggest event in human history.
Unfortunately, it might also be the last,
unless we learn how to avoid the risks.

Stephen Hawking

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 data set containing the phase constituents of 55 microstructures as well as corresponding

This chapter is based on the article: **W. Li**, M. Vittorietti, G. Jongbloed, J. Sietsma, *Microstructure–property relation and machine learning prediction of hole expansion capacity of high-strength steels*, Journal of Materials Science **56**, 19228 (2021).

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.

keywords

microstructure constituents, hole expansion capacity, statistical analysis, machine learning

4.1. 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. Fig. 4.1 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

$$\lambda = \frac{D_h - D_0}{D_0}. \tag{4.1}$$

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] reveals the damage evolution on both the punched and machined edges, which indicates the possibility of building predictive models based on physical understanding.

Table 4.1 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 physics-based explanation [2, 6, 19].

The information obtained from Table 4.1 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

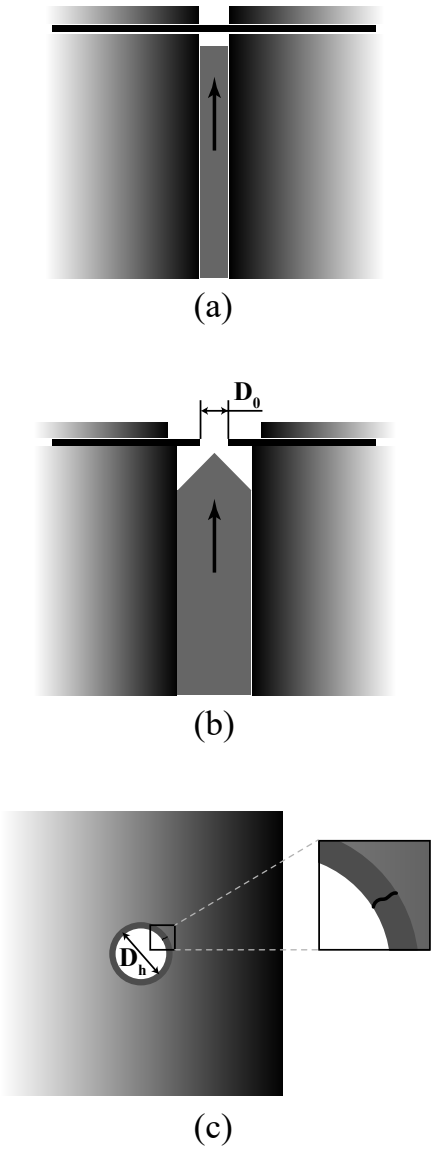


Figure 4.1: Schematic illustration of HEC test after ISO 16630. The standard prescribes $D_0 = 10\,\text{mm}$.

Table 4.1: 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 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]

Table 4.1 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 with the HEC values are collected, as shown in Table 4.A.1 (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 4.1. 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.

4.2. Data analysis

As shown in Table 4.A.1 (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 of $15\,\%$ on HEC values for martensitic steels [23].

4.2.1. 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. 4.2. Fig. 4.2(a) and (b) show the scatter plot of martensite (without tempered martensite) and ferrite fraction in relation to hole expansion capacity, while Fig. 4.2(c) 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. 4.2(b) is a linear fitting of all data points of ferrite volume fraction and hole expansion capacity. All curved lines in Fig. 4.2 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. 4.2(a), 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. 4.2(b), 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.A.1, that consists of a large volume fraction of ferrite and secondary phase pearlite. An opposite trend to martensite is shown in Fig. 4.2(c) when looking into the relation between the total bainite volume fraction and HEC. High HEC values are found only between 30% and 40% bainite. Fig. 4.2(d) 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. 4.2(d). The lower volume fraction of retained austenite shows higher HEC than the group of higher volume fraction. In Fig. 4.2(d) the bainite fractions are also indicated for the microstructures. Relating

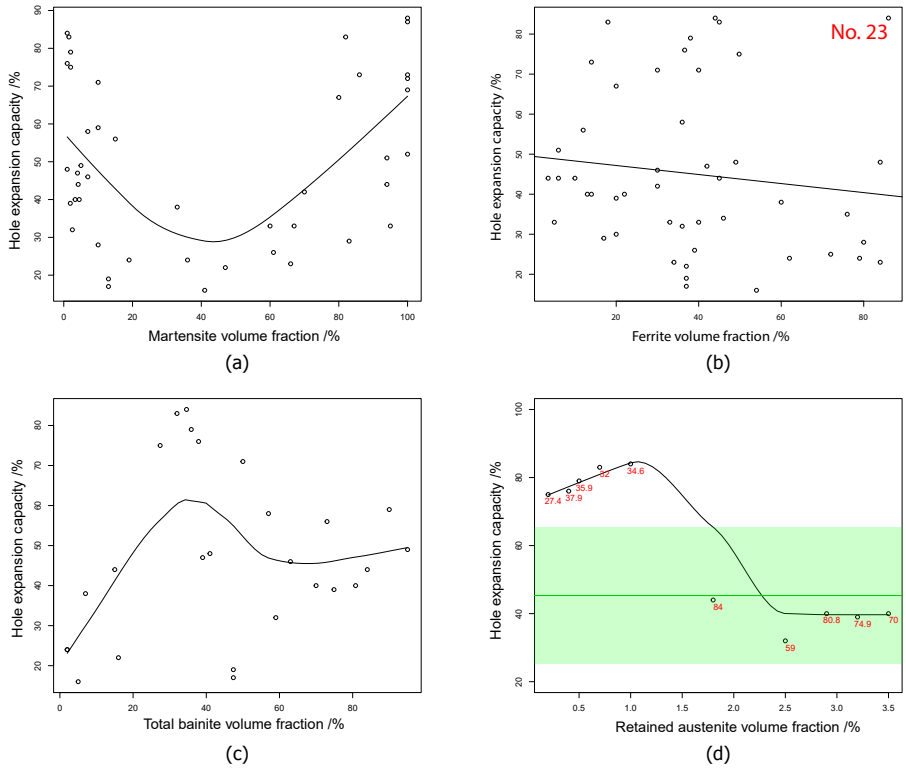


Figure 4.2: HEC relation with martensite (a), ferrite (b), total bainite (c) and retained austenite (d) volume fractions. The numbers in (d) given with the data points are the total bainite fraction in percent. The green line shows the average HEC value, while the green shaded region shows the standard deviation, of the microstructures without RA.

these values to Fig. 4.2(c), 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. 4.2(d), at the level of the values for 2-4% RA. The present data therefore does 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, λ = 48%) and No. 29 (10% P, 80% F, 10% M, λ = 28%). The difference between these two HEC values is therefore primarily the result of the difference in martensite and pearlite fractions. The reduction from λ = 48% for 1% martensite to λ = 28% for 10% martensite is stronger than the general trend in Fig. 4.2(a), which points at a positive effect of pearlite on the HEC.

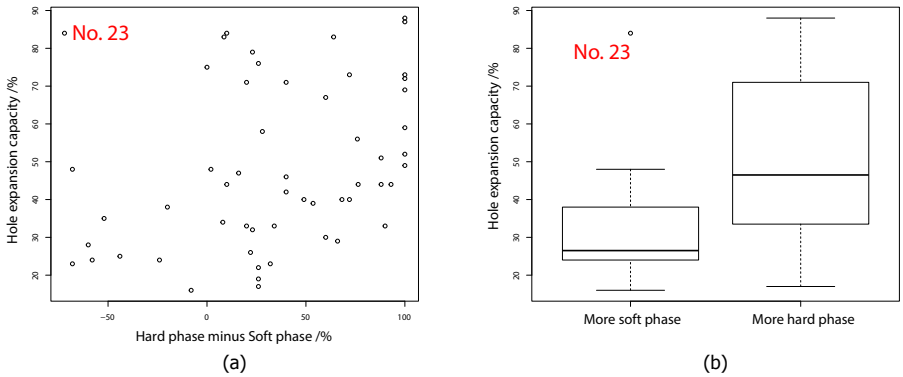


Figure 4.3: HEC relation with difference between hard and soft phase fractions (a) and HEC difference for two groups with respect to the volume fraction of hard and soft phase (b).

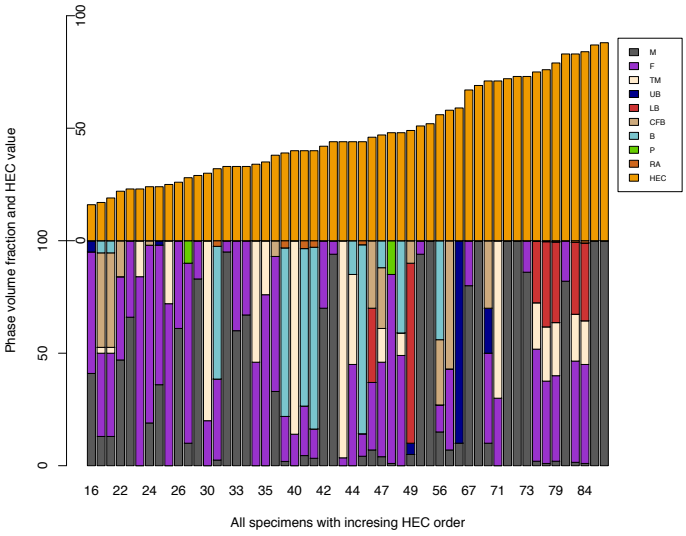
Difference between volume fraction of phases

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. 4.3(a). The scatter plot shows an increase of HEC when the hard phase volume fraction is increasing. When the data set is divided into two groups, as the boxplot in Fig. 4.3(b) 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. 4.2(b).

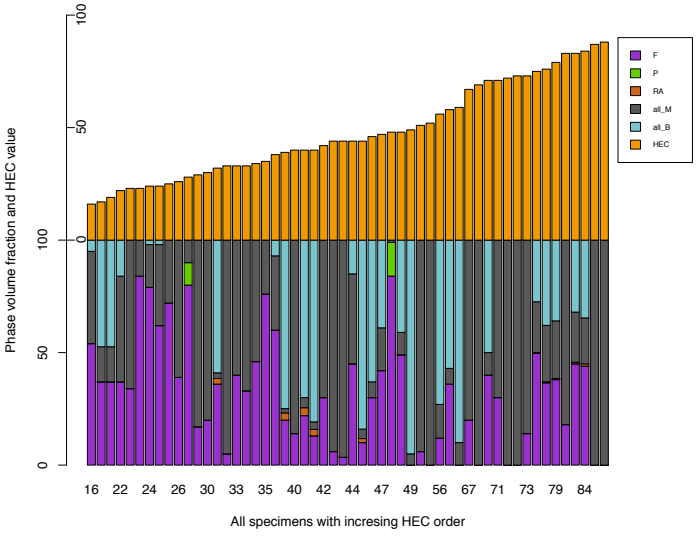
The one exception mentioned in Section 4.2.1, No. 23 in Table 4.A.1, is also marked in Fig. 4.3. It clearly shows that this No. 23 sample is an outlier with exceptionally high HEC while containing more soft phase, which is considered to be an artefact of the testing procedure. Hence in the following statistical analysis, this No. 23 sample is deleted from the dataset.

4.2.2. 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. 4.4 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.



(a)



(b)

Figure 4.4: Phase composition with increasing HEC order, (a): all the individual phase are present, (b): combining all the martensite together and all the bainite together.

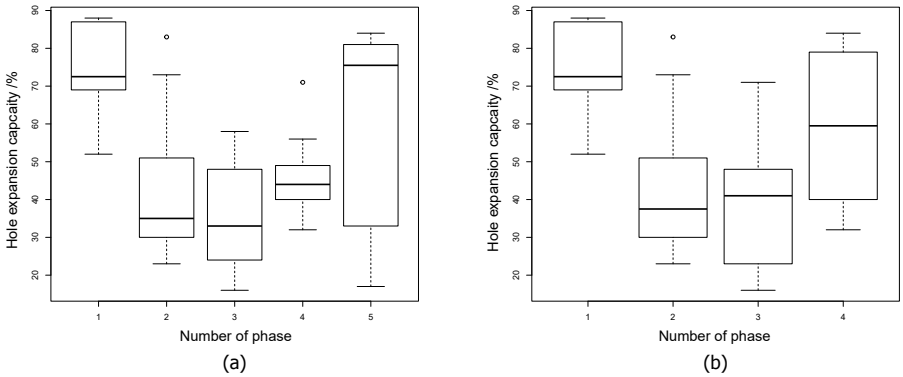


Figure 4.5: HEC relation with number of phases (a) and with number of phases while combining all bainite and all martensite (b).

4.2.3. HEC in relation to number of phases

Fig. 4.5 shows the box-plot 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 all martensite. In Table 4.2, the t-value is the estimate (2nd column, the coefficient for each input variable) divided by its standard error (3nd 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.

Fig. 4.5 shows that the one-phase category has the highest HEC values, while in Table 4.2, the p-values (last column) for one, two, three, four phases in non-combed phase fractions (a) and one, two and three phases in combining all bainite and all martensite condition (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 phases or four phases, as shown in Fig. 4.5 and Table 4.2. Only the five phases without combining and four phases with combining have increased HEC, since these structures belong to the ones mentioned in Section 4.2.2 which have the combination of ferrite, martensite and bainite with the volume ratio around

Table 4.2: 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. 4.5.

(a) Number of phases	Estimate	Std. Error	t value	P
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**
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

2:1:1. These microstructures all have a low volume fraction of retained austenite and both martensite and tempered martensite are present.

4.2.4. 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 Section 4.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, hence leads to wrong estimates of the coefficients [26]. In Fig. 4.6, 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 on Fig. 4.6 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 seen in Fig. 4.2(a). 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 show the effect for 0-50% martensite, which is therefore the same as the trend seen in Fig. 4.2(a).

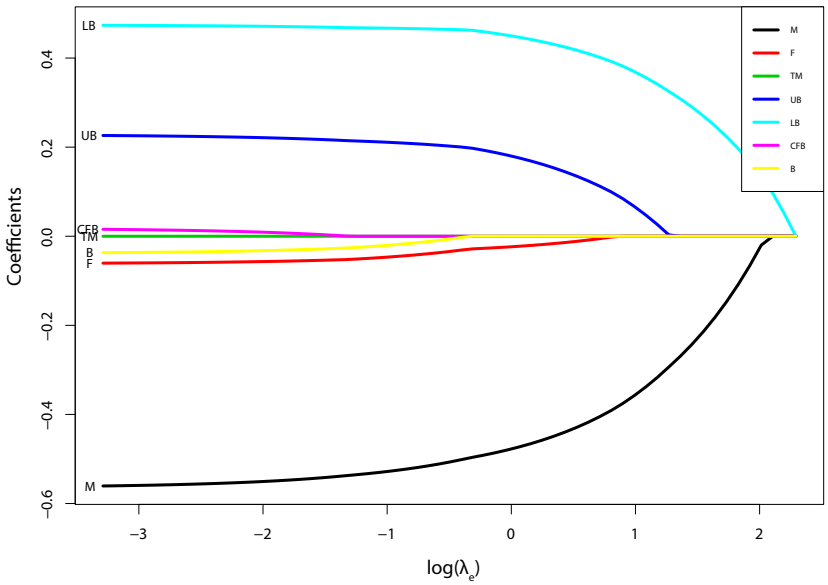


Figure 4.6: LASSO plot on the influence of different phase fractions on HEC.

4.3. 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.A.1, we selected five different machine learning methods:

1. Linear regression (lm)
2. Linear regression with Elastic Net regularization (glmnet)
3. Conditional Inference Tree regression (ctree2)
4. Random Forest regression (cforest)
5. Deep learning (keras)

Detailed information on these methods can be found in appendix Section 4.B. The first four methods are applied using the 'caret' library [37], adopted in the R [24] environment. For the first four methods, 10-fold 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, 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 step 1-4 10 times (10-fold cross validation) and calculate the mean performance, i.e. the average RMSE on both training and testing data over 10 repeated runs.

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

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\lambda_{\mathrm{p},i} - \lambda_{\mathrm{r},i})^2}{N}}. \tag{4.2}$$

4.3.1. Machine learning model performance

The performance of all five machine learning models is shown in Fig. 4.7. The two linear regression methods (1&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 $\pm 16\,\%$. Comparing to 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 $\pm 9\,\%$, 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. 4.8, 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 of the experimental data quality and increase of quantity of the data, the authors believe that the prediction accuracy can be further enhanced.

4.3.2. Machine learning model interpretation

The Conditional Inference Tree regression model and Random Forest regression model both give rise to a ranking of importance of the independent variables, which is shown in Fig. 4.9. 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 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

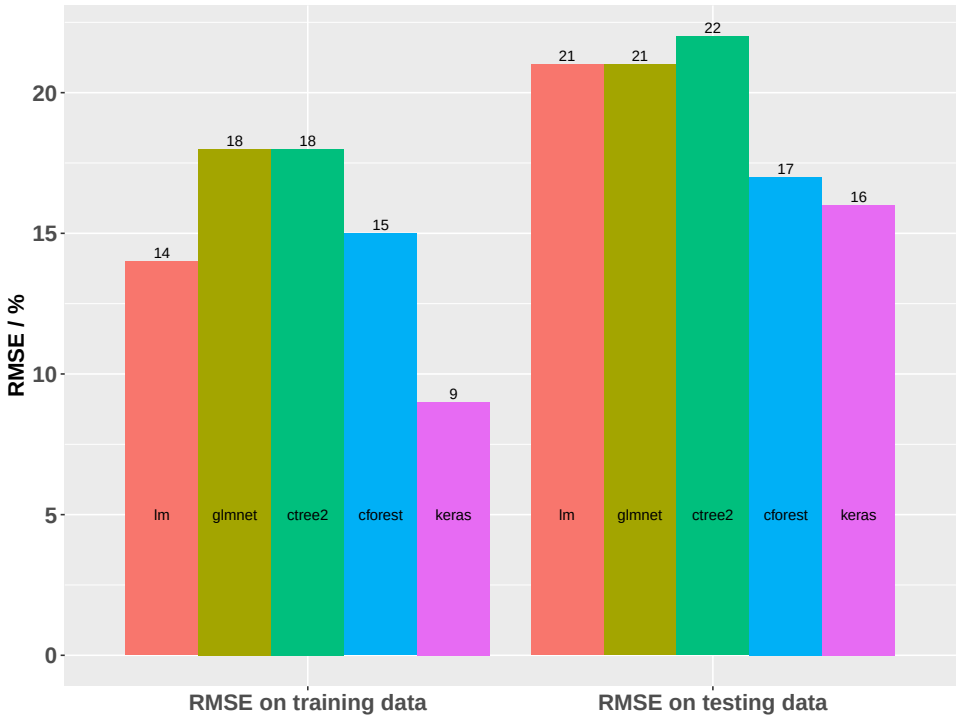


Figure 4.7: Performance comparison of all five machine learning methods.

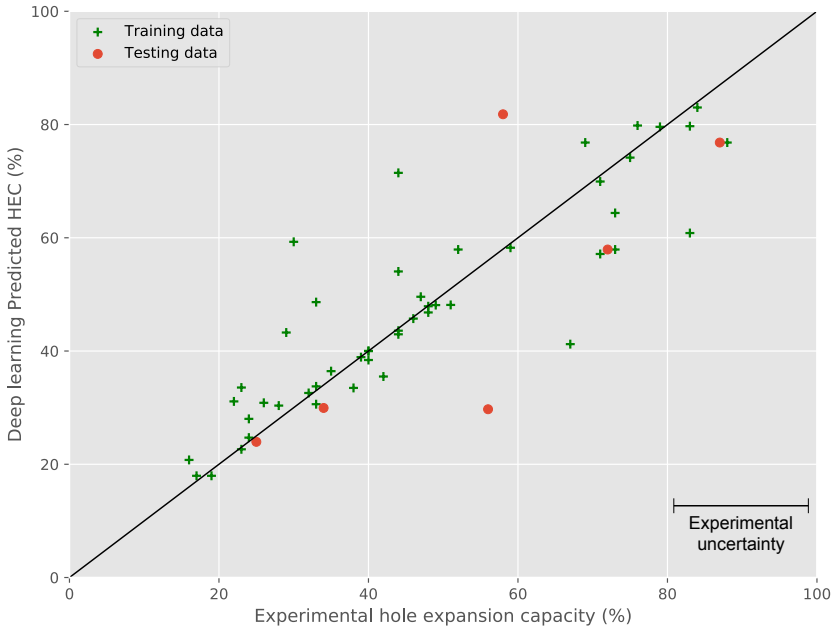


Figure 4.8: Deep learning prediction on hole expansion capacity.

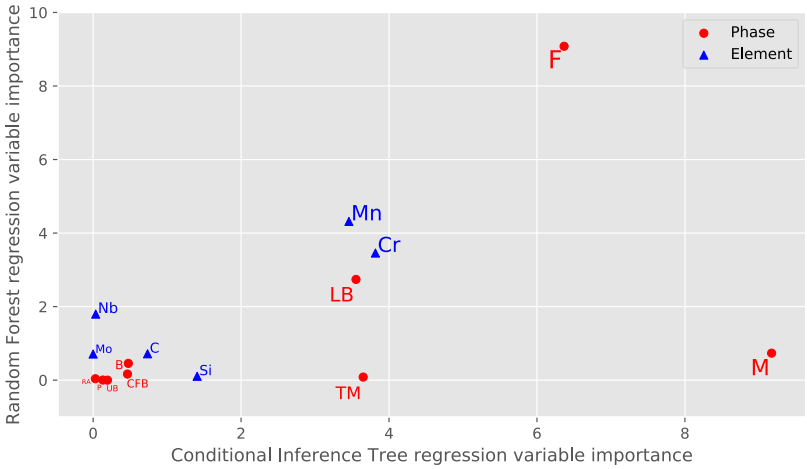


Figure 4.9: Variable importance plot for both conditional inference tree regression model and random forest regression model. Phases are depicted in blue, chemical elements in red.

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. 4.10. At each node of the decision tree, one specific input variable is selected, according to algorithms mentioned in appendix Section 4.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, it is false for 6 samples with an average HEC of 74%, as shown in node 16. The samples of each of these nodes are further seperated 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 changing of HEC with different variables is not monotonic. Table 4.3 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. 4.10 classify the highest HEC with either fully martensitic structure or the combination of martensite, lower bainite and ferrite. Meanwhile, node 8 in Fig. 4.10 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%.

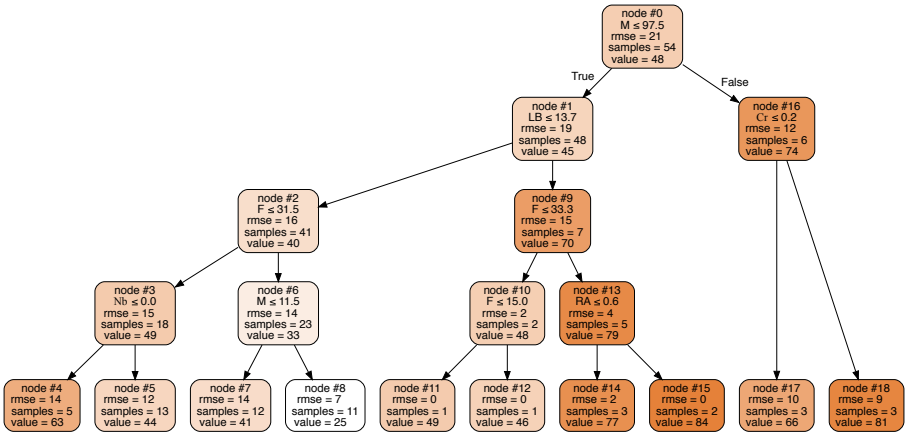


Figure 4.10: The conditional inference tree regression model plot, phase fractions in percentage and chemical content in weight percent. Colors indicate the magnitude of the average HEC in the node.

Table 4.3: Summary of the conditional inference tree based on the range of HEC values.

No.	HEC (%)	Phase fraction (%) and chemical content (wt%)	node
I	25	$11.5 < f_{\mathrm{M}} \leqslant 97.5$; $f_{\mathrm{LB}} \leqslant 13.7$; $f_{\mathrm{F}} > 31.5$	node 8
II	40-50	$f_{\mathrm{M}} \leqslant 97.5$; $f_{\mathrm{LB}} \leqslant 13.7$; $f_{\mathrm{F}} \leqslant 31.5$; $c_{\mathrm{Nb}} > 0$ $f_{\mathrm{M}} \leqslant 11.5$; $f_{\mathrm{LB}} \leqslant 13.7$; $f_{\mathrm{F}} > 31.5$ $f_{\mathrm{M}} \leqslant 97.5$; $f_{\mathrm{LB}} > 13.7$; $f_{\mathrm{F}} \leqslant 33.3$	node 5 node 7 node 10
III	60-70	$f_{\mathrm{M}} \leqslant 97.5$; $f_{\mathrm{LB}} \leqslant 13.7$; $f_{\mathrm{F}} \leqslant 31.5$; $c_{\mathrm{Nb}} = 0$ $f_{\mathrm{M}} > 97.5$; $c_{\mathrm{Cr}} \leqslant 0.2$	node 14 node 17
IV	over 70	$f_{\mathrm{M}} \leqslant 97.5$; $f_{\mathrm{LB}} > 13.7$; $f_{\mathrm{F}} > 33.3$ $f_{\mathrm{M}} > 97.5$; $c_{\mathrm{Cr}} > 0.2$	node 13 node 18

4.4. Discussion

Comparing the summary in Table 4.1 and the statistical analysis in Section 4.2.1, there are some clearly contradictory results. The summary in Table 4.1 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 4.1. 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 Section 4.2, in whichever way the data is 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 vally 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 of 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 $\pm 16\,\%$ 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 data sets. The authors believe that significant improvement of 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 morphol-

ogy, which are normally considered to have distinct impact on mechanical behavior, this study shows valuable results with limited materials information.

4.5. Conclusions

This study focuses on data acquired from 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 tend 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.

References

[1] ISO 16630:2017, *Metallic materials – Sheet and strip – Hole expanding test*, Standard (International Organization for Standardization, 2017).

[2] A. Karelova, C. Krempaszky, E. Werner, P. Tsipouridis, T. Hebesberger, and A. Pichler, *Hole Expansion of Dual-phase and Complex-phase AHS Steels - Effect of Edge Conditions,* steel research international **80**, 71 (2009).

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

[4] L. Xu, F. Barlat, M. G. Lee, K. S. Choi, and X. Sun, *Hole expansion of dual phase steels,* in *WIT Transactions on the Built Environment*, Vol. 124 (2012) pp. 75–83.

[5] R. Comstock, D. Scherrer, and R. Adamczyk, *Hole Expansion in a Variety of Sheet Steels,* Journal of Materials Engineering and Performance **15**, 675 (2006).

[6] Y.-R. Cho, J.-H. Chung, H.-H. Ku, and I.-B. Kim, *Effect of controlled cooling on the formability of TS 590 MPa grade hot-rolled high strength steels,* Metals and Materials **5**, 571 (1999).

[7] X. Fang, Z. Fan, B. Ralph, P. Evans, and R. Underhill, *The relationships between tensile properties and hole expansion property of C-Mn steels,* Journal of Materials Science **38**, 3877 (2003).

[8] X. Fang, Z. Fan, B. Ralph, P. Evans, and R. Underhill, *Effects of tempering temperature on tensile and hole expansion properties of a C–Mn steel,* Journal of Materials Processing Technology **132**, 215 (2003).

[9] D. Casellas, A. Lara, D. Frómeta, D. Gutiérrez, S. Molas, L. Pérez, J. Rehrl, and C. Suppan, *Fracture Toughness to Understand Stretch-Flangeability and Edge Cracking Resistance in AHSS,* Metallurgical and Materials Transactions A **48**, 86 (2017).

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

[11] D. Frómeta, M. Tedesco, J. Calvo, A.
, S. Molas, and D. Casellas, *Assessing edge cracking resistance in AHSS automotive parts by the Essential Work of Fracture methodology,* Journal of Physics: Conference Series **896**, 012102 (2017).

[12] D. Frómeta, S. Parareda, A. Lara, S. Molas, D. Casellas, P. Jonsén, and J. Calvo, *Identification of fracture toughness parameters to understand the fracture resistance of advanced high strength sheet steels,* Engineering Fracture Mechanics **229**, 106949 (2020).

[13] K. Ismail, A. Perlade, P. J. Jacques, and T. Pardoen, *Outstanding cracking resistance of fibrous dual phase steels,* Acta Materialia **207**, 116700 (2021).

[14] A. Bareggi, E. Maire, O. Bouaziz, and M. Di Michiel, *Damage in dual phase steels and its constituents studied by X-ray tomography,* International Journal of Fracture **174**, 217 (2012).

[15] T. de Geus, J. van Duuren, R. Peerlings, and M. Geers, *Fracture initiation in multi-phase materials: A statistical characterization of microstructural damage sites,* Materials Science and Engineering: A **673**, 551 (2016).

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

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

[18] M. Kahziz, T. F. Morgeneyer, M. Mazière, L. Helfen, O. Bouaziz, and E. Maire, *In situ 3D Synchrotron Laminography Assessment of Edge Fracture in Dual-Phase Steels: Quantitative and Numerical Analysis,* Experimental Mechanics **56**, 177 (2016).

[19] K. Hasegawa, K. Kawamura, T. Urabe, and Y. Hosoya, *Effects of Microstructure on Stretch-flange-formability of 980 MPa Grade Cold-rolled Ultra High Strength Steel Sheets,* ISIJ International **44**, 603 (2004).

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

[21] K.-i. Sugimoto, J. Sakaguchi, T. Iida, and T. Kashima, *Stretch-flangeability of a High-strength TRIP Type Bainitic Sheet Steel.* ISIJ International **40**, 920 (2000).

[22] X. Chen, H. Jiang, Z. Cui, C. Lian, and C. Lu, *Hole Expansion Characteristics of Ultra High Strength Steels,* Procedia Engineering **81**, 718 (2014).

[23] E. Atzema, M. Borsutzki, M. Braun, S. Brockmann, M. Buelter, B. Carlsson, P. Larour, and A. Richter, *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* (2012) pp. 171–184.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2013).

[25] W. Cleveland, E. Grosse, and W. Shyu, *Local regression models,* in *Statistical models in S*, edited by J.M. Chambers and T.J. Hastie (Wadsworth & Brooks/Cole, 1992) Chap. 8.

[26] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer Texts in Statistics, Vol. 103 (Springer New York, New York, NY, 2013).

[27] C. Heumann, M. Schomaker, and Shalabh, *Introduction to Statistics and Data Analysis* (Springer International Publishing, Cham, 2016).

[28] A. Ghatak, *Machine Learning with R* (Springer Singapore, Singapore, 2017).

[29] W. Li, M. Vittorietti, G. Jongbloed, and J. Sietsma, *The combined influence of grain size distribution and dislocation density on hardness of interstitial free steel,* Journal of Materials Science & Technology **45**, 35 (2020).

[30] J. Wei, X. Chu, X. Sun, K. Xu, H. Deng, J. Chen, Z. Wei, and M. Lei, *Machine learning in materials science,* InfoMat **1**, 338 (2019).

[31] P. Santak and G. Conduit, *Predicting physical properties of alkanes with neural networks,* Fluid Phase Equilibria **501**, 112259 (2019).

[32] R. Jose and S. Ramakrishna, *Materials 4.0: Materials big data enabled materials discovery,* Applied Materials Today **10**, 127 (2018).

[33] M. A. Bessa, P. Glowacki, and M. Houlder, *Bayesian Machine Learning in Metamaterial Design: Fragile Becomes Supercompressible,* Advanced Materials , 1904845 (2019).

[34] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, *Machine-learning-assisted materials discovery using failed experiments,* Nature **533**, 73 (2016), arXiv:NIHMS150003 .

[35] D. M. Dimiduk, E. A. Holm, and S. R. Niezgoda, *Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering,* Integrating Materials and Manufacturing Innovation **7**, 157 (2018).

[36] E. A. Holm, *In defense of the black box,* Science **364**, 26 (2019).

[37] M. Kuhn, *Building predictive models in r using the caret package,* Journal of Statistical Software, Articles **28**, 1 (2008).

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

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

[40] A. F. Agarap, *Deep learning using rectified linear units (relu),* (2018).

[41] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization,* (2014).

[42] R. V. Hogg, A. T. Craig, and J. W. McKean, *Introduction to mathematical statistics* (Pearson Education, Boston, Mass, 2014).

[43] H. Guo, P. Zhou, A.-m. Zhao, C. Zhi, R. Ding, and J.-x. Wang, *Effects of Mn and Cr contents on microstructures and mechanical properties of low temperature bainitic steel,* Journal of Iron and Steel Research International **24**, 290 (2017).

[44] K.-i. Sugimoto, A. Kanda, R. Kikuchi, S.-i. Hashimoto, T. Kashima, and S. Ikeda, *Ductility and Formability of Newly Developed High Strength Low Alloy TRIP-aided Sheet Steels with Annealed Martensite Matrix.* ISIJ International **42**, 910 (2002).

[45] C. Landron, O. Bouaziz, E. Maire, and J. Adrien, *Characterization and modeling of void nucleation by interface decohesion in dual phase steels,* Scripta Materialia **63**, 973 (2010).

[46] S. Allain and O. Bouaziz, *Microstructure based modeling for the mechanical behavior of ferrite–pearlite steels suitable to capture isotropic and kinematic hardening,* Materials Science and Engineering: A **496**, 329 (2008).

[47] H. Zou and T. Hastie, *Addendum: Regularization and variable selection via the elastic net,* Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**, 768 (2005).

[48] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York, New York, NY, 2009).

[49] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees* (Chapman and Hall/CRC, 1984).

[50] J. R. Quinlan, *C4.5 : programs for machine learning* (Morgan Kaufmann Publishers, San Mateo, Calif., 1993).

[51] T. Hothorn, K. Hornik, and A. Zeileis, *Unbiased Recursive Partitioning: A Conditional Inference Framework,* Journal of Computational and Graphical Statistics **15**, 651 (2006).

[52] H. Strasser, H. Strasser, and C. Weber, *On the Asymptotic Theory of Permutation Statistics,* MATHEMATICAL METHODS OF STATISTICS **2** (1999).

[53] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, *Bias in random forest variable importance measures: Illustrations, sources and a solution,* BMC Bioinformatics **8**, 25 (2007).

[54] S. Haykin and S. S. Haykin, *Neural Networks and Learning Machines* (Prentice Hall, 2009).

Appendix

4.A. 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 is shown in Table 4.A.1.

Table 4.A.1: Raw data from the literature 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	0.110	2.10	0.15	0	0.20	0.020
2	67	80.0	20.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
3	88	100.0	0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
4	73	100.0	0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
5	51	94.0	6.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
6	87	100.0	0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
7	72	100.0	0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
8	33	95.0	5.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
9	69	100.0	0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
10	42	70.0	30.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
11	33	60.0	40.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
12	83	82.0	18.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
13	33	67.0	33.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
14	26	61.0	39.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
15	73	86.0	14.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
16	29	83.0	17.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
17	23	66.0	34.0	0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
18	44	94.0	6.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
19	38	33.0	60.0	0	0	0	7	0	0	0	0.110	2.10	0.15	0	0.20	0.020
20	24	19.0	79.0	0	0	0	2	0	0	0	0.135	1.50	0.50	0	0	0.015

Table 4.A.1: (continued)

No.	HEC	M	F	TM	UB	LB	CFB	B	P	RA	C	Mn	Si	Cr	Mo	Nb
21	22	47.0	37.0	0	0	0	16	0	0	0	0.140	1.70	0.18	0.3	0.15	0
22	24	36.0	62.0	0	2	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
23	84	0	86.0	0	0	0	0	0	14	0	0.135	1.50	0.50	0	0	0.015
24	16	41.0	54.0	0	5	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
25	71	10.0	40.0	0	20	0	30	0	0	0	0.110	2.10	0.15	0	0.20	0.020
26	48	1.0	84.0	0	0	0	0	0	15	0	0.135	1.50	0.50	0	0	0.015
27	59	10.0	0	0	90	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
28	46	7.0	30.0	0	0	33.0	30	0	0	0	0.110	2.10	0.15	0	0.20	0.020
29	28	10.0	80.0	0	0	0	0	0	10	0	0.135	1.50	0.50	0	0	0.015
30	49	5.0	0	0	5	80.0	10	0	0	0	0.140	1.70	0.18	0.3	0.15	0
31	19	13.0	37.0	2.6	0	0	42	5.4	0	0	0.170	1.70	0.18	0.3	0.15	0
32	17	13.0	37.0	2.6	0	0	42	5.4	0	0	0.170	1.70	0.18	0.3	0.15	0
33	56	15.0	12.0	0	0	0	29	44.0	0	0	0.170	1.70	0.18	0.3	0.16	0
34	47	4.0	42.0	15.0	0	0	27	12.0	0	0	0.110	2.10	0.15	0	0.20	0.020
35	58	7.0	36.0	0	0	0	57	0	0	0	0.140	2.10	0.20	0.3	0	0.020
36	35	0	76.0	24.0	0	0	0	0	0	0	0.144	1.52	0.48	0	0	0.018
37	30	0	20.0	80.0	0	0	0	0	0	0	0.172	1.59	0.51	0	0	0.017
38	71	0	30.0	70.0	0	0	0	0	0	0	0.170	1.63	0.47	0	0	0.017
39	44	0	3.5	96.5	0	0	0	0	0	0	0.178	1.41	0.44	0	0	0.016
40	40	0	14.0	86.0	0	0	0	0	0	0	0.135	1.53	0.18	0	0	0.015
41	25	0	72.0	28.0	0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
42	34	0	46.0	54.0	0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
43	23	0	84.0	16.0	0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
44	44	0	45.0	40.0	0	0	0	15.0	0	0	0.130	1.00	0.20	0	0	0.015
45	48	0	49.0	10.0	0	0	0	41.0	0	0	0.130	1.00	0.20	0	0	0.015
46	75	2.0	49.8	20.6	0	27.4	0	0	0	0.2	0.140	2.10	0.20	0.3	0	0.020

Table 4.A.1: (continued)

No.	HEC	M	F	TM	UB	LB	CFB	B	P	RA	C	Mn	Si	Cr	Mo	Nb
47	84	1.0	44.0	19.4	0	34.6	0	0	0	1.0	0.140	2.10	0.20	0.3	0	0.020
48	83	1.5	45.0	20.8	0	32.0	0	0	0	0.7	0.140	2.10	0.20	0.3	0	0.020
49	76	1.0	36.6	24.1	0	37.9	0	0	0	0.4	0.140	2.10	0.20	0.3	0	0.020
50	79	2.0	38.0	23.6	0	35.9	0	0	0	0.5	0.140	2.10	0.20	0.3	0	0.020
51	32	2.5	36.0	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	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	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	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	74.9	0	3.2	0.170	1.70	0.18	0.3	0.15	0.015

4.B. Principles of regression methods

4.B.1. Linear regression with Elastic Net regularization (glmnet)

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

$$\hat{f}_{\text{ElasticNet}}(\mathbf{x}) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_j, \tag{4.B.1}$$

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:

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

Here n is the number of the data points, x_{ij} is the ith observation corresponding to the jth variable, 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.

4.B.2. Conditional Inference Tree regression (ctree2)

A decision tree is a model in the form of a tree structure, which breaks the data set 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 selection bias towards 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.

4.B.3. 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 a Conditional Random Forest which can be simply seen as averaging multiple Conditional Inference Tree results [48, 51, 53].

4.B.4. 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. Fig. 4.B.1 shows an example of the neural network which uses all 15 microstructure features to predict the hole expansion capacity. The first 15 input neurons build up the input layer, while the output layer, in this case, is just one neuron, i.e. Hole Expansion Capacity. The hidden layers are in the middle. Fig. 4.B.1 shows two hidden layers, while in the actual case, two hidden layers of 100 and 30 neurons are applied. This kind of fully connected neural network is called multilayer perceptron. Data flows from the input layer through the hidden layers and finally arrives output layer. The mathematics for calculating the value Y of each neuron from the neurons in the previous layer is [48, 54]:

$$Y = F(B + \sum_{i=1}^{n} w_i x_i), \tag{4.B.3}$$

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, *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.

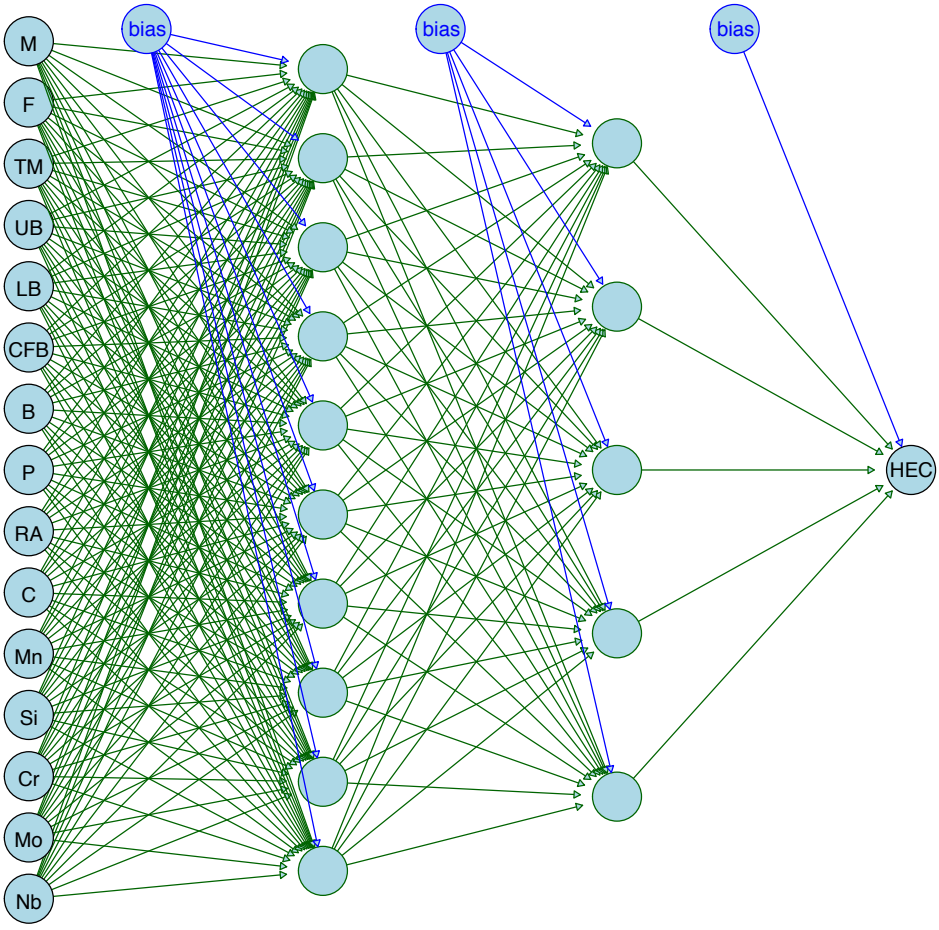


Figure 4.B.1: An example of the neural network with the input variables and output variable used in Section 4.3.

5

Machine learning prediction of bendability and the corresponding microstructure-property relation analysis of high strength steels

Notice that the stiffest tree is most easily cracked,
while the bamboo or willow survives by bending with the wind.

Bruce Lee

Bendability, while being an important factor affecting the application of advanced high strength steels in various industries, is not yet well explained in terms of relationship with microstructure features, due to the lack of overview on various phases and the lack of data containing both microstructure features and bendability. In this study, an experimental data set containing the phase volume fractions and the chemical composition of the microstructure as well as corresponding property, i.e. bendability, is collected from the literature. Statistical analysis on these data is conducted with the focus on bendability in relation to individual phases, combinations of phases and number of phases. In addition, different machine learning methods are

applied to predict bendability based on both phase fractions and chemical composition. The results show that the tree-based models give the best prediction of bendability. Meanwhile, the influence of different microstructure features on bendability is revealed by quantifying and comparing the different effects.

keywords

bendability, microstructure constituents, statistical analysis, machine learning

5.1. Introduction

Advanced high strength steels are widely adopted in different industries, such as automotive, aerospace and construction [1, 2]. With the complex shapes being needed in the specific user cases, high standards of forming properties are required, such as bendability, which describes the ability of materials to undergo bending force without cracking initiation along the bending line [3]. As shown in Fig. 5.1, the three point bending test according to ISO7438 standard [4] is carried out with a punch forcing the sheet material onto two rollers. The thickness of the testing materials should be less than 25 mm. The distance between the two rollers should be three times the thickness of the test materials plus the diameter/thickness of the punch. The punch stops at the point where the crack initiates along the bending line. In order to ensure the plain strain condition throughout the bending process, the width and length of the sheet should be at least 20 times the sheet thickness. A better bendability is shown by a smaller minimum bending radius, which is normally characterized by the inner radius R_i. To compare the bendability for materials with different thickness, the minimum bending radius is usually divided by the thickness of the material t, hence leads to the ratio R_i/t [5]. A lower R_i/t corresponds to a better bendability.

Various studies have shown that different materials features, such as phases, precipitates and surface roughness, have influence on the bendability [6–11]. However, due to the lack of data points with bendability and corresponding microstructure features, and the lack of a general view over multiple phases, these results are not sufficient to fully understand the bendability in relation to microstructure features. In order to study further the relations between bendability and microstructure features, 44 sets of data containing the volume fraction of phases and chemical composition as well as the bendability are collected, as shown in Table 5.A.1, from an RFCS final report [12]. Since the original work did not focus on the specific relation between bendability and microstructure features, this chapter will look in detail into this relation and apply machine learning algorithms to predict bendability from phase volume fractions and chemical compositions. In addition, the importance of the different microstructure features are shown and discussed.

5.2. Data analysis

As shown in Table 5.A.1 (see Appendix), the obtained dataset contains data from 44 low-alloy steel specimens [12]. For each specimen, the minimum bending ratio (bendability) R_i/t with phase fractions in volume percentage and chemical composition in weight percentage are collected. The phases are martensite (M), ferrite (F), tempered martensite (TM), upper bainite (UB), lower bainite (LB), carbide-free bainite (CFB), bainite (B) and pearlite (P). The present study is based on this collected dataset.

5.2.1. Bendability in relation to volume fraction of phases

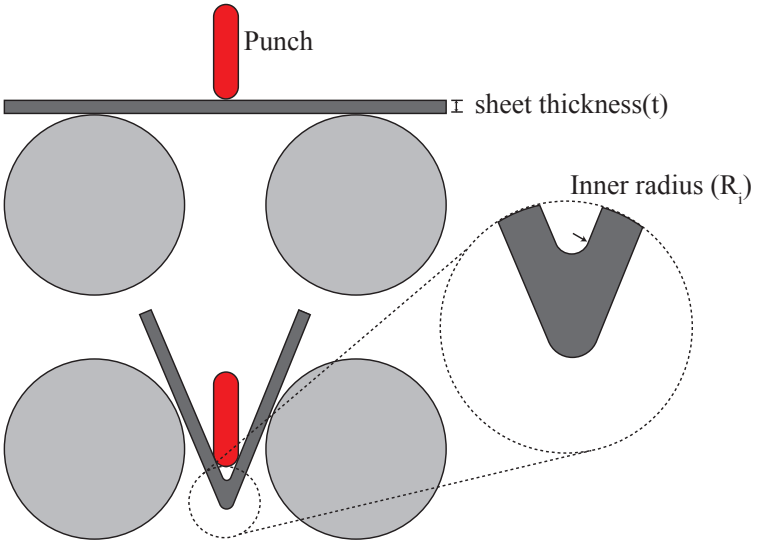


Figure 5.1: Schematic drawing of bending test.

Volume fraction of individual phases

Based on the obtained data in Table 5.A.1 (see Appendix), the individual influence of phase fractions on the bendability is shown in Fig. 5.2 with the scatter plot of both martensite and ferrite volume fraction with the bendability. The straight lines show the linear fitting of all data points between martensite and ferrite fractions in relation to the bendability. The legend used on each sample point represents the most abundant phase except martensite in Fig. 5.2(a) and ferrite in Fig. 5.2(b).

As shown in Fig. 5.2, there is clear influence from both martensite and ferrite on bendability. Martensite has a negative influence. With higher volume fraction of martensite, R_{i}/t increases, i.e. the bendability becomes worse. Meanwhile, the ferrite volume fraction has the opposite effect. But since the scatter plot shows strong deviations from the linear regression line, this only gives the trend of the influence. As shown in Fig. 5.2(a), all specimens which have more than 65% volume fraction of martensite have bendability value large than 3, which indicates the poor bendability for those samples. Meanwhile for ferrite in Fig. 5.2(b), those with more than 50% volume fraction of ferrite have good bendability.

While looking at the most abundant phase except martensite in Fig. 5.2(a), most of the samples have ferrite as the most abundant phase. Among the other cases, two samples with lower bainite as the most abundant phase except martensite have better bendability than other samples with upper bainite, bainite or carbide-free bainite as the most abundant phase except martensite. This could be associated with the carbide formation within bainitic ferrite in the lower bainite microstructure. While looking at the most abundant phase except ferrite in Fig. 5.2(b), half of the samples have martensite as the most abundant phase except ferrite. In

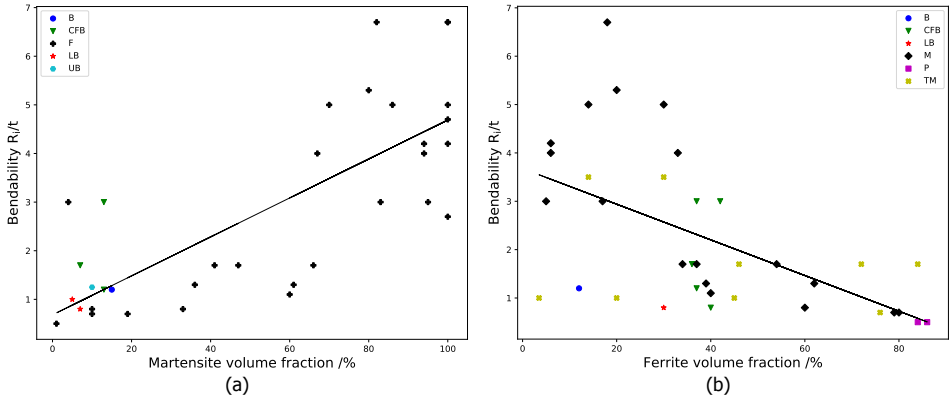


Figure 5.2: Bendability relation with martensite (a) and ferrite (b) volume fraction. The legend represents the most abundant phase except martensite in (a) (or ferrite in (b)) for each sample.

those cases, only a few of them, which have large fraction of ferrite present (more than 40%), have good bendability. This again indicates the negative effect of the martensite phase. Among the other cases in Fig. 5.2(b) that the samples with tempered martensite as the most abundant phase except ferrite generally have good bendability, while there are two exceptions, which could be caused by the degree of tempering of these specific samples on which the original dataset does not give information. Meanwhile, samples with pearlite as the most abundant phase except ferrite also exhibit good bendability. These three cases are all associated with more than 80% volume fraction of ferrite. Combining other cases located at the lower right corner of Fig. 5.2(b), it is shown that samples with around 80% volume fraction of ferrite and either pearlite, martensite or tempered martensite as the secondary phase show excellent bendability. It is generally well-known that ferrite phase has good ductility but low strength. It is interesting to see in Fig. 5.2(b), that the samples on the lower left corner with presence of tempered martensite as the most abundant phase except ferrite have good bendability. Comparing to those good bendability samples with higher volume fraction of ferrite, these cases have better strength performance, which could be beneficial if not only the bendability is required, but also the strength of the samples is a requirement.

Difference between volume fraction of phases

Due to the limited number of specimens which contain bainite and pearlite, there is no clear relation directly associated to bendability, therefore the phases are sorted into two groups, i.e. soft phase and hard phase. Here except ferrite, all other phases shown are assumed to be hard phase, including martensite, bainite and pearlite. The relationship between bendability and the soft and hard phases is shown in Fig. 5.2 (b) and Fig. 5.3. When separating the data points into two groups, i.e. more soft phase and more hard phase, a clear distinction between these two groups is shown in the box plot in Fig. 5.3. The group containing more soft phase (the left box) has a mean value of the minimum bending ratio R_i/t below

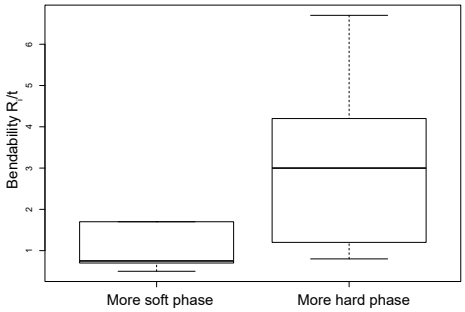


Figure 5.3: Bendability relation with hard and soft phases.

1, while the group containing more hard phase (the right box) has a mean value of R_i/t over 3. This indicates the significant beneficial effect of soft phase on the bendability.

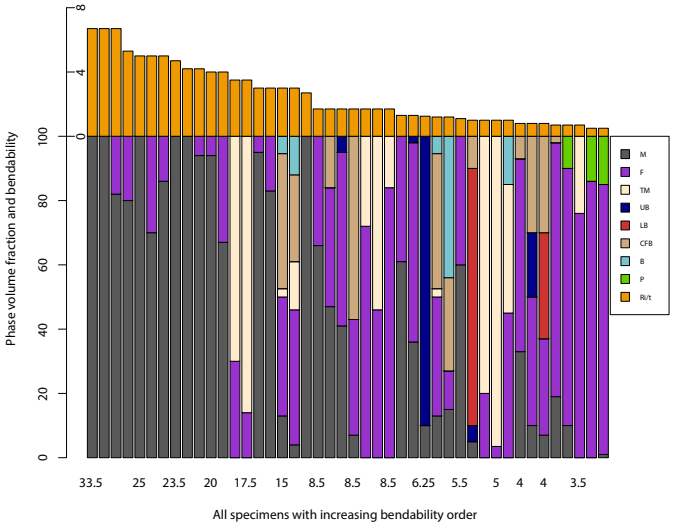
5.2.2. Bendability in relation to combinations of phases

In order to check if the different phase volume fractions have combined effects on the bendability, the phase compositions are plotted with a stacked bar plot in the order of increasing bendability, as shown in Fig. 5.4. While Fig. 5.4 (a) shows all individual phases, Fig. 5.4 (b) has all the martensite and all the bainite combined.

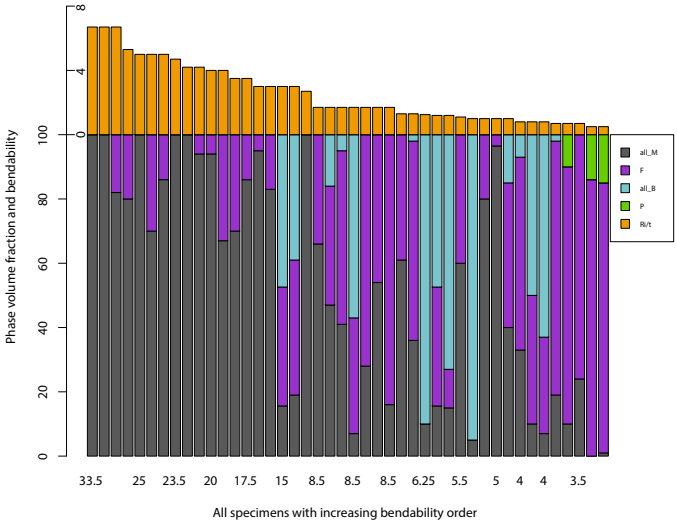
It is clearly shown that the ferritic matrix has beneficial effect on the bendability, while martensitic structure has negative effect. When ferrite is present with volume fraction of more than 50%, the presence of the secondary phases makes significant changes to the bendability. Among those secondary phases, pearlite has a significant positive contribution to bendability comparing to martensite and bainite. But since there are only three samples with pearlite present in the whole dataset, the influence of the fraction of pearlite on bendability is hard to reveal here. The presence of a similar fraction of tempered martensite does not always show good bendability. This could be caused by the degree of tempering in the processing of the microstructure. Combination of ferrite and pearlite, combination of ferrite and martensite, and combination of these three phases all contribute to better bendability. In these good combinations, ferrite has around 80% volume fraction.

5.2.3. Bendability in relation to number of phases

Fig. 5.5 shows the box-plot of bendability in relation to number of phases. The number shown above the median line is the average ferrite volume fraction in the corresponding group. In order to quantify this relationship, we introduce a linear model with number of phases as the explanatory variable and bendability as the response variable, the relationship is shown in Table 5.1. The number of phases has either 5 or 3 categories, corresponding to the non-combined phase fractions



(a)



(b)

Figure 5.4: Phase composition with increasing Bendability order, (a): all the individual phase are present, (b): combining all the martensite together and all the bainite together.

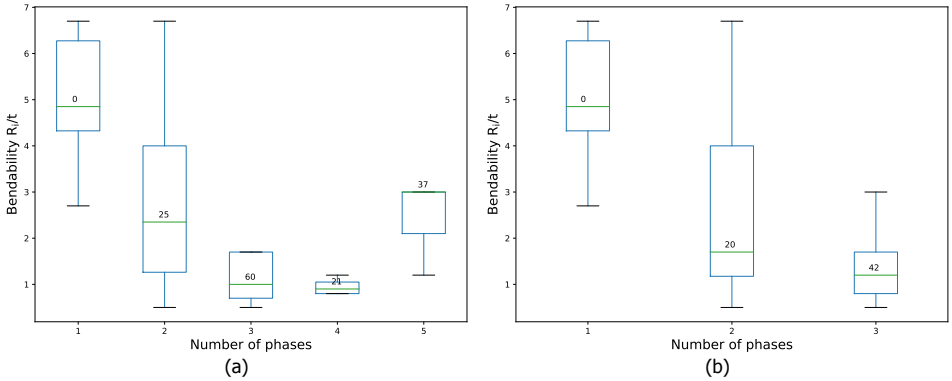


Figure 5.5: Bendability relation with number of phases, without combination (a) and with combination of all martensite and all bainite (b). The number shows above the median line is the average ferrite volume fraction in the corresponding group.

5

or the case of combining all bainite and all martensite. Detailed explanation of the t-value and p-value can be found in Section 4.2.3. Typically a p-value less than 0.05 indicates that there is a relation between the explanatory variable and the response variable. Since all p-values in Table 5.1 are distinctly smaller than 0.05, it is obvious that with the increasing of number of phases, the bendability ratio decreases, hence better bendability. This shows that complex-phase steels generally have better bendability. In the case of decreasing bendability with 5 phases present in Fig. 5.5, it is because in this case, large faction of ferritic matrix was replaced with martensite and bainite, which have negative a effect on the bendability.

Looking into the numbers above the median line in Fig. 5.5, which represent the average ferrite volume fraction of the corresponding group, it can be seen that the influence from the number of phases cannot be explained only by the change of volume fraction of ferrite in the case of separating all bainite and all martensite in Fig. 5.5(a). In the case of combination of all martensite and all bainite in Fig. 5.5(b). The influence of the number of phases follows the same trend as the influence of volume fraction of ferrite, i.e. the higher the volume fraction of ferrite is, the better the bendability is.

5.3. Prediction of bendability with both phase fraction and chemical content

With the powerful prediction performance shown by machine learning applied in the materials science field [13–19] and in order to predict the bendability based on microstructure features gathered in Table 5.A.1, five machine learning methods are chosen to be applied in this case:

1. Linear regression (lm)

2. Linear regression with Elastic Net regularization (glmnet)

Table 5.1: Results summary for linear regression between Number of phases and bendability for both separated phases (a) and with combination of all martensite and all bainite (b), corresponding to Fig. 5.5.

(a) Number of phases	Estimate	Std. Error	t value	P
1 (Intercept)	5.0	0.59	8.47	0.0000
2	-2.2	0.67	-3.35	0.0018
3	-3.9	0.76	-5.09	0.0000
4	-4.1	0.93	-4.34	0.0001
5	-2.6	1.02	-2.54	0.0151
(b) Number of phases	**Estimate**	**Std. Error**	**t value**	**P**
1 (Intercept)	5.0	0.60	8.32	0.0000
2	-2.3	0.67	-3.43	0.0014
3	-3.7	0.71	-5.15	0.0000

3. Conditional Inference Tree regression (ctree2)
4. Random Forest regression (cforest)
5. Deep Learning (keras)

A similar procedure as shown below and the applied libraries ('caret' in R and 'keras' in Python) are described in the hole expansion capacity Section 4.3 and the machine learning methods details can be found in Section 4.B. In short, the procedure is as follows.

1. Data partitioning into training and testing set (random: 90% of the data in training dataset, 10% of the data in testing dataset),
2. Feed training data to train the model,
3. Predict testing target (Bendability) using the trained model,
4. Calculate the performance (calculate RMSE on both training and testing data),
5. Repeat process 1-4 for 10 times for randomly chosen training and testing dataset (10-fold cross validation) and calculate the mean performance.

5.3.1. Machine learning model performance

The performance of the machine learning models is shown in Fig. 5.6. All five models have similar prediction accuracy on the training dataset. But on the prediction performance of the testing dataset, the two linear models (lm and glmnet) and the deep learning model have worse performance comparing to the two tree-based models (ctree2 and cforest). With the two tree based models, the accuracies of the prediction of bendability R_i/t are both ± 1.

5.3.2. Machine learning model interpretation

Based on the two tree-based models, the combined ranking of importance is shown in Fig. 5.7, where the zoomed-in plot shows the detailed information from the lower left corner. The Conditional Inference Tree's feature importance is calculated by the sum of the reduction of variance to the parent node weighted by the probability of reaching that node that is caused by the certain feature. Higher value indicates high importance. Random Forest takes the feature importance from each tree and averages the importance to obtain the rank of importance of all features. It is clearly shown that the martensite fraction plays the most important role in determining the bendability. Ferrite ranks second. This shows the significant influence of hard phase, especially martensite, on the bendability. The zoomed-in plot shows that the other phase fractions and chemical content have significantly less influence on bendability based on the two tree-based prediction models.

Fig. 5.8 shows a tree built from the conditional inference tree model. It can be seen that from the top node 0, martensite volume fraction already has huge impact on the bendability. The two nodes below, node 1 and node 8, show the mean value of bendability of 1.4 and 4.7, respectively, by the separation criteria of martensite volume fraction of 66%. Further looking down the tree, the best bendability shows up with tempered martensite volume fraction below 1%, ferrite volume fraction above 70% and hence martensite volume fraction below 30%. This corresponds to the findings in Section 5.2.2. The worst bendability shows up with martensite volume fraction between 66% and 84% in combination of more than 0.14 wt% of carbon content. If considering the bendability of 1.3 as the threshold, nodes 3, 4 and 6 represent the microstructures with good bendability. Those microstructures all contain less than 66% martensite. Nodes 3 and 4 have less than 1% tempered martensite (no tempered martensite present in these cases), while node 6 has more than 1% of tempered martensite and less than 1.5 wt% Mn.

5.4. Discussion

Considering the data analysis done in the previous sections, it seems that prediction of bendability based on the phase volume fractions and chemical content is not a simple task. While only knowing the phase volume fractions and chemical content, it is hard to get a precise prediction of bendability R_{i}/t, comparing to the previous prediction on hole expansion capacity in Chapter 4. The prediction accuracy of the Conditional Inference Tree regression and Random Forest regression are both ± 1, which only gives a relatively large range prediction. There are multiple reasons for this limitation of prediction. One could be the limited number of data points. Giving more data for the training of the model, it could have better performance, especially for the deep learning model. The other could be the reason that only phase fractions and chemical content cannot describe the bendability adequately. Since it can be seen from Fig. 5.4 that even for some similar combination of phase fractions, there is still significant variations of the experimental bendability. Besides phase volume fractions and chemical content, there would be many other influencing factors on bendability, such as grain size, grain morphology, texture and even surface condition. This is the limitation of the current used dataset.

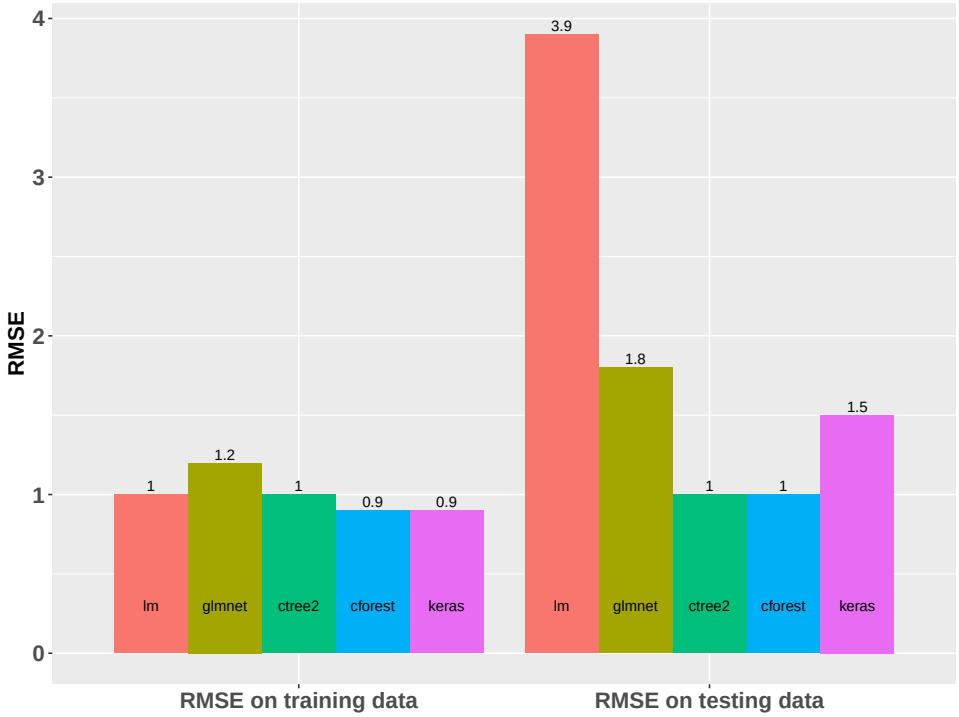


Figure 5.6: Performance comparison of all five machine learning models on the prediction of bendability.

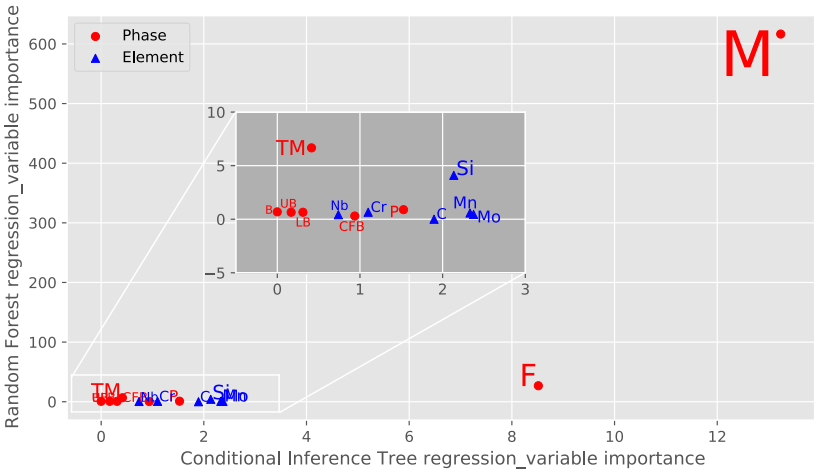


Figure 5.7: Variable importance plot for both conditional inference tree regression model and random forest regression model on bendability.

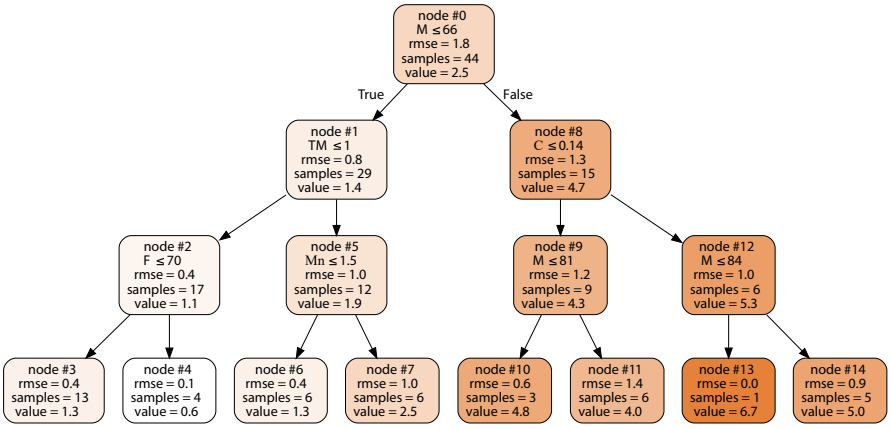


Figure 5.8: The conditional inference tree regression model plot on prediction of bendability, phase fractions in percentage and chemical content in weight percent. Colors indicate the magnitude of the average bendability in the node.

While the prediction modelling gives the most important variables to be martensite and ferrite, it justifies the findings in Section 5.2. Generally, the soft phase ferrite contributes positively to bendability while the hard phase martensite has the opposite effect. This is highly related to the ductility of the phases. Increasing strength often leads to the degradation of ductility. Bendability is closely related to both material strength and ductility. On the one hand, better ductility leads to smaller inner bending radius, hence better bendability. On the other hand, higher strength of the material only increases the force needed for bending, while it does not increase the bendability.

5.5. Conclusions

This study concentrates on investigation of the relationship among the data collected from literature about microstructure features with corresponding bendability values. The following conclusions can be drawn from the research.

- Prediction of bendability only based on phase volume fractions and chemical content cannot be done as accurately as experimental determination. The prediction accuracy is ± 1 on R_i/t.
- Based both on the prediction models and the data analysis, martensite and ferrite are recognized to be the most important microstructure features for bendability among phase fractions and chemical content.
- Soft phase mainly contributes positively to bendability, while hard phase contributes negatively.

- Complex-phase steels generally have better bendability, especially under certain combination of phase, such as combination of ferrite and pearlite, combination of ferrite and martensite, and combination of these three phase, with the condition of ferrite being the matrix phase and have volume fraction above 70%.

References

[1] T. K. Roy, B. Bhattacharya, C. Ghosh, and S. Ajmani, *Advanced high strength steel* (Springer, 2018).

[2] H. Ban and G. Shi, *A review of research on high-strength steel structures,* Proceedings of the Institution of Civil Engineers - Structures and Buildings **171**, 625 (2018).

[3] S. Heibel, T. Dettinger, W. Nester, T. Clausmeyer, and A. Tekkaya, *Damage Mechanisms and Mechanical Properties of High-Strength Multiphase Steels,* Materials **11**, 761 (2018).

[4] ISO 7438:2020, *Metallic materials – Bend test*, Standard (International Organization for Standardization, 2020).

[5] C. Horvath, *Advanced steels for lightweight automotive structures,* in *Materials, Design and Manufacturing for Lightweight Vehicles* (Elsevier, 2010) pp. 35–78.

[6] M. Dao and M. Li, *A micromechanics study on strain-localization-induced fracture initiation in bending using crystal plasticity models,* Philosophical Magazine A **81**, 1997 (2001).

[7] K.-i. SUGIMOTO, B. YU, Y.-i. MUKAI, and S. IKEDA, *Microstructure and Formability of Aluminum Bearing TRIP-Aided Steels with Annealed Martensite Matrix,* ISIJ International **45**, 1194 (2005).

[8] A. J. Kaijalainen, P. Suikkanen, L. P. Karjalainen, and J. J. Jonas, *Effect of Austenite Pancaking on the Microstructure, Texture, and Bendability of an Ultrahigh-Strength Strip Steel,* Metallurgical and Materials Transactions A **45**, 1273 (2014).

[9] A. Kaijalainen, P. Suikkanen, L. Karjalainen, and D. Porter, *Influence of subsurface microstructure on the bendability of ultrahigh-strength strip steel,* Materials Science and Engineering: A **654**, 151 (2016).

[10] A. Kaijalainen, V. Kesti, L. Troive, A.-M. Arola, T. Liimatainen, M. Hemmilä, J. Kömi, and D. Porter, *Superior bendability of direct-quenched 960 MPa strip steels,* Procedia Manufacturing **15**, 676 (2018).

[11] C. Soyarslan, M. Malekipour Gharbi, and A. Tekkaya, *A combined experimental – numerical investigation of ductile fracture in bending of a class of*

ferritic–martensitic steel, International Journal of Solids and Structures **49**, 1608 (2012).

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

[13] J. Wei, X. Chu, X. Sun, K. Xu, H. Deng, J. Chen, Z. Wei, and M. Lei, *Machine learning in materials science,* InfoMat **1**, 338 (2019).

[14] P. Santak and G. Conduit, *Predicting physical properties of alkanes with neural networks,* Fluid Phase Equilibria **501**, 112259 (2019).

[15] R. Jose and S. Ramakrishna, *Materials 4.0: Materials big data enabled materials discovery,* Applied Materials Today **10**, 127 (2018).

[16] M. A. Bessa, P. Glowacki, and M. Houlder, *Bayesian Machine Learning in Metamaterial Design: Fragile Becomes Supercompressible,* Advanced Materials , 1904845 (2019).

[17] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, *Machine-learning-assisted materials discovery using failed experiments,* Nature **533**, 73 (2016), arXiv:NIHMS150003 .

[18] D. M. Dimiduk, E. A. Holm, and S. R. Niezgoda, *Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering,* Integrating Materials and Manufacturing Innovation **7**, 157 (2018).

[19] E. A. Holm, *In defense of the black box,* Science **364**, 26 (2019).

Appendix

5.A. Raw data from literature

The collected and organized data from the work of Ryde *et al.* about the bendability with chemical composition and phase fractions is shown in Table 5.A.1.

Table 5.A.1: Raw data from the literature with bendability with phase fractions in percentage and chemical content in wt%, the phases are martensite (M), ferrite (F), tempered martensite (TM), upper bainite (UB), lower bainite (LB), carbide-free bainite (CFB), bainite (B) and pearlite (P).

No.	R_i/t	M	F	TM	UB	LB	CFB	B	P	C	Mn	Si	Cr	Mo	Nb
1	6.7	100.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
2	5.3	80.0	20.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
3	6.7	100.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
4	4.7	100.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
5	4.0	94.0	6.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
6	5.0	100.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
7	2.7	100.0	0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
8	3.0	95.0	5.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
9	4.2	100.0	0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
10	5.0	70.0	30.0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
11	1.1	60.0	40.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
12	6.7	82.0	18.0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
13	4.0	67.0	33.0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
14	1.3	61.0	39.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
15	5.0	86.0	14.0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
16	3.0	83.0	17.0	0	0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
17	1.7	66.0	34.0	0	0	0	0	0	0	0.135	1.50	0.50	0	0	0.015
18	4.2	94.0	6.0	0	0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
19	0.8	33.0	60.0	0	0	0	7.0	0	0	0.110	2.10	0.15	0	0.20	0.020
20	0.7	19.0	79.0	0	0	0	2.0	0	0	0.135	1.50	0.50	0	0	0.015

Table 5.A.1: (continued)

No.	R_i/t	M	F	TM	UB	LB	CFB	B	P	C	Mn	Si	Cr	Mo	Nb
21	1.7	47.0	37.0	0	0	0	16.0	0	0	0.140	1.70	0.18	0.3	0.15	0
22	1.3	36.0	62.0	0	2.0	0	0	0	0	0.110	2.10	0.15	0	0.20	0.020
23	0.5	0	86.0	0	0	0	0	0	14.0	0.135	1.50	0.50	0	0	0.015
24	1.7	41.0	54.0	0	5.0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
25	0.8	10.0	40.0	0	20.0	0	30.0	0	0	0.110	2.10	0.15	0	0.20	0.020
26	0.5	1.0	84.0	0	0	0	0	0	15.0	0.135	1.50	0.50	0	0	0.015
27	1.2	10.0	0	0	90.0	0	0	0	0	0.140	1.70	0.18	0.3	0.15	0
28	0.8	7.0	30.0	0	0	33.0	30.0	0	0	0.110	2.10	0.15	0	0.20	0.020
29	0.7	10.0	80.0	0	0	0	0	0	10.0	0.135	1.50	0.50	0	0	0.015
30	1.0	5.0	0	0	5.0	80.0	10.0	0	0	0.140	1.70	0.18	0.3	0.15	0
31	1.2	13.0	37.0	2.6	0	0	42.0	5.4	0	0.170	1.70	0.18	0.3	0.15	0
32	3.0	13.0	37.0	2.6	0	0	42.0	5.4	0	0.170	1.70	0.18	0.3	0.15	0
33	1.2	15.0	12.0	0	0	0	29.0	44.0	0	0.170	1.70	0.18	0.3	0.16	0
34	3.0	4.0	42.0	15.0	0	0	27.0	12.0	0	0.110	2.10	0.15	0	0.20	0.020
35	1.7	7.0	36.0	0	0	0	57.0	0	0	0.140	2.10	0.20	0.3	0	0.020
36	0.7	0	76.0	24.0	0	0	0	0	0	0.144	1.52	0.48	0	0	0.018
37	1.0	0	20.0	80.0	0	0	0	0	0	0.172	1.59	0.51	0	0	0.017
38	3.5	0	30.0	70.0	0	0	0	0	0	0.170	1.63	0.47	0	0	0.017
39	1.0	0	3.5	96.5	0	0	0	0	0	0.178	1.41	0.44	0	0	0.016
40	3.5	0	14.0	86.0	0	0	0	0	0	0.135	1.53	0.18	0	0	0.015
41	1.7	0	72.0	28.0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
42	1.7	0	46.0	54.0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
43	1.7	0	84.0	16.0	0	0	0	0	0	0.125	1.52	0.20	0	0	0.015
44	1.0	0	45.0	40.0	0	0	0	15.0	0	0.130	1.00	0.20	0	0	0.015

III

Deformation to Fracture

6

Relationship between deformation properties and fracture properties

We are like islands in the sea,
separate on the surface but connected in the deep.

William James

The relationship between the deformation properties and the fracture properties is an interesting topic. Considering the cost and the complication of the testing process for fracture properties, it is valuable to evaluate the relationship between these two kinds of properties in order to predict fracture properties based on deformation properties. This chapter focuses on experimental data which contains both types of mechanical properties. The study on the relationships between Charpy impact energy and common tensile test properties is covered in this chapter. The focus of this chapter is not only on establishing the relation but also on the prediction of the Charpy impact energy from tensile properties using machine learning models and the interpretation of the model.

Keywords

Charpy impact energy, tensile properties, random forest, Partial Dependence Plot, Accumulated Local Effects Plot, SHapley Additive exPlanations

6.1. Introduction

Deformation properties and fracture properties are two kinds of important mechanical properties of steels. The first two parts of this thesis study the influence of microstructure features on selected mechanical properties, both deformation related and fracture related. The third part of this thesis is focusing on the relationship between deformation properties and fracture properties. This should give insight into these complicated relations and hence deepen the understanding of the mechanical behavior of steels. Although the relation between deformation properties and fracture properties has not yet been extensively studied, there are studies showing relations between specific properties. Various studies [1–4] summarized the relation between yield strength and Charpy impact energy for various kinds of steels, where yield strength is generally negatively correlated to Charpy impact energy. The fracture properties are generally controlled by the inherent resistance to crack initiation and growth [1, 5]. Interestingly in the work done by Liao *et al.* [6] on Mg alloys, the Charpy impact energy is enhanced when the mean grain size is decreased under 3 µm, and in the meantime, yield strength is increased with the decreasing of mean grain size. Moreover, Cao *et al.* [7] achieved ultrahigh Charpy impact energy with high tensile strength with ferrite/martensite laminated steels. These contradictory results show the complicated relations between Charpy impact energy and the tensile properties.

In order to have better understanding of the relations, a large number of experimental data is required. Due to the complicated process of gathering complete datasets by doing individual mechanical tests, publicly available datasets are valuable. There are various kinds of materials data sources available both in the literature and from online databases, which contain microstructure features and mechanical properties [8].

For the study of the relation between deformation properties and fracture properties, a systematically obtained dataset is required. The dataset used in this chapter was collected by Dr. Gareth Conduit, who is now working at the University of Cambridge, from various publicly available data sources [9]. The whole dataset contains the chemical composition and various mechanical properties (yield strength (YS), ultimate tensile strength (UTS), uniform elongation (UE), Charpy impact energy (K_{V}) and fracture toughness (K_{Ic})) of 821 Advanced High-Strength Steels (AHSS), which is now available on the website https://citrination.com [9].

In this chapter, the related deformation and fracture properties are briefly introduced in Section 6.2, followed by the relation analysis between properties in Section 6.3 based on the obtained dataset. The prediction of Charpy impact energy is described in Section 6.4 and the interpretation of the machine learning model is explained in detail in Section 6.5.

6.2. Deformation and fracture properties

6.2.1. Tensile properties

Tensile test is the most commonly used mechanical testing method to characterize the mechanical properties of materials. In general, the sample is subjected to

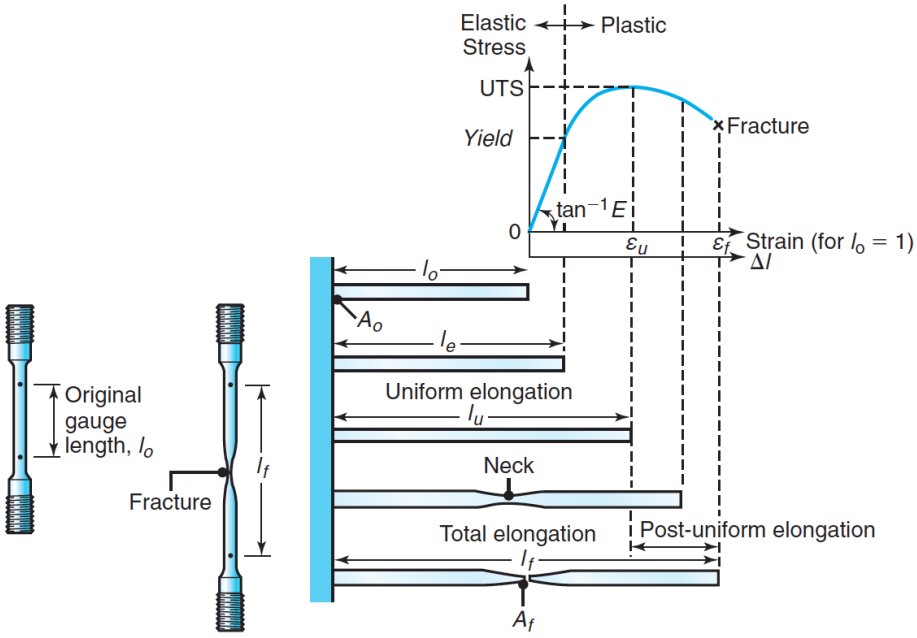


Figure 6.1: Tensile test specimen before and after pulling (left) and tensile test process (right) (Adapted from [10]).

controlled strain until fracture. A typical method to characterize the tensile behavior is through stress-strain curves as schematically shown in Fig. 6.1. The engineering stress σ is defined as the ratio between applied force F and original cross-section area A_0, given by $\sigma = F/A_0$. The engineering strain ϵ is defined as the ratio between elongation and original gauge length l_0, given by $\epsilon = (l - l_0)/l_0$, where l is the instantaneous length of the gauge.

Fig. 6.1 shows a schematic engineering stress-strain curve. As the loading increases, materials first behave elastically. The engineering stress-strain curve shows an approximately linear region. The linear relationship in the first region of the stress-strain curve is known as Hooke's law [11]. The ratio between elastic stress and strain is the modulus of elasticity, also known as Young's modulus, $E = \sigma/\epsilon$, which is the slope of the curve in the elastic region. In practice, the yield strength σ_y is hard to determine due to the gradual transition from the elastic region to the plastic region. By drawing a line parallel to the elastic region at the offset strain of 0.2%, the yield strength is defined as the stress at the intersection of the offset line with the engineering stress-strain curve [10, 12]. The yield strength given in the database can be assumed to be determined by this method. Once the yield strength is reached, plastic deformation starts. Until reaching the ultimate tensile strength σ_u (UTS), which marks the highest engineering stress during a tensile test, the elongation is uniform. At the UTS necking starts and deformation becomes local. Young's modulus, YS and UTS are all strength-related properties,

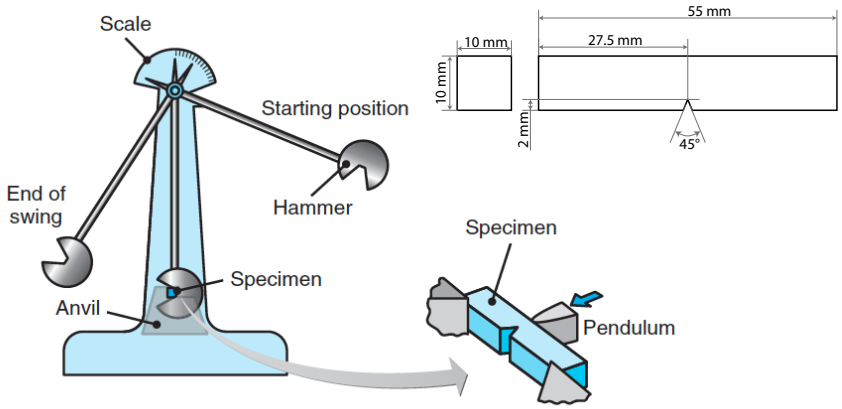


Figure 6.2: Charpy test set up and specimen dimensions (adapted from [10]).

while UE and total elongation indicate the ductility of the material. The area under the stress-strain curve $\int \sigma \mathrm{d}\epsilon$ in principle represents the toughness of the material.

6

6.2.2. Charpy impact energy

The most commonly used method to determine toughness is the Charpy test, due to its easy set-up and simple experimental procedure. The Charpy test set-up and specimen can be seen in Fig. 6.2. The specimen is supported at both ends on the platform. The hammer is released from a fixed position and breaks the specimen at the notch, which requires energy. Depending on the rising height of the hammer after breaking the notched specimen, the consumed impact energy (K_{V}, since the notch is V-shaped) can be calculated. Due to the dependence of the impact energy on the specimen dimension, materials are tested under standard specimen dimensions, i.e. the standard full-size Charpy test specimen CVN (10 mm $\times$ 10 mm $\times$ 55 mm), and the toughness is given by the energy required to break a sample of these dimensions.

6.3. Relations between properties

Looking from the angle of energy absorption for both tensile test and Charpy test, there are certain correlations between them. During the Charpy test, the specimen deforms at high strain rate until fracture. Hence the measured Charpy impact energy is the sum of energy absorbed in both deformation process and fracture process. Comparing to the tensile test, by only recording the yield strength, ultimate tensile strength and uniform elongation, only the deformation energy is taken into consideration, where deformation occurs in a simpler loading mode than in a Charpy test and at a lower stain rate. Since the actual fracture process is not included in the tensile properties, the relation with the Charpy impact energy is expected to be only partial.

The dataset used in this study mainly focuses on austenitic stainless steels with addition of Cr, Ni and Mo. The complete dataset after cleaning is shown in appendix Table 6.A.1. The cleaning of the raw dataset is done by removing the entries with missing values in tensile properties or Charpy impact energy. The cleaned dataset contains 263 entries, which have full information on tensile test properties and Charpy impact energy data. Since the yield strength and ultimate tensile strength are both in the range of 1000 MPa to 2100 MPa and the uniform elongation is in the range of 4% to 24%, this dataset only represents advanced high strength steels.

In order to identify the relation between tensile properties and Charpy impact energy, first the pair plot is applied on the complete dataset as shown in Fig. 6.3. Here the area under the tensile curve (A_d) is calculated from yield strength (σ_y), ultimate tensile strength (σ_u) and uniform elongation (ϵ_u), with the equation $A_\mathrm{d} = \sigma_\mathrm{y} * \epsilon_\mathrm{u} + 0.5 * (\sigma_\mathrm{u} - \sigma_\mathrm{y}) * \epsilon_\mathrm{u}$. A_d is thus approximated by the sum of a rectangle (side lengths yield strength and uniform elongation) and a right triangle (side lengths uniform elongation and difference between yield strength and ultimate tensile strength). The plots on the diagonal are the distributions of the corresponding property, while the other plots are the relation plots between corresponding properties on each row and column. The distribution plots of UTS and YS are schematically similar while the distribution plots of ϵ_u and A_d are also similar. However K_V shows a much different distribution. These can be directly connected to the relation between corresponding properties. There is a clear positive relation between yield strength and ultimate tensile strength. Both yield strength and ultimate tensile strength have negative relation with uniform elongation. Such relations are well known. When considering the fourth row, the relation of tensile properties with A_d, only the uniform elongation shows pronounced positive effect on A_d. The focus in this study is on the four plots on the bottom row, shows the relations between Charpy impact energy and each tensile property. For these the trends are not so clear. Delimiting lines are qualitatively indicated on the bottom row of Fig. 6.3, where all data points are located in a triangle, which indicates an upper limit for Charpy impact energy at certain values for yield strength, ultimate tensile strength and uniform elongation. This indicates that Charpy impact energy is highly strength related.

Fig. 6.4 shows a colored scatter plot with corresponding contour plot for UTS and uniform elongation, where the color indicates the value for Charpy impact energy. There is a trend that data points are located near the diagonal line, which follows the common rule that the increase of strength often implies a decrease of the ductility, as was already seen in the same plot in Fig. 6.3. More interestingly, it can be seen from the contour plot that Charpy impact energy generally increases with increasing ductility, but with exceptions at the low UTS and high UE ranges. When the uniform elongation is over 20%, it shows a decrease of Charpy impact energy with further increasing uniform elongation.

Fig. 6.5 shows the relation between deformation energy and Charpy impact energy (on a log-scale). Here the deformation energy is calculated from A_d, which yields the deformation energy per unit volume, and the established deformation volume in a Charpy sample. It is assumed that the deformed volume of the Charpy

Figure 6.3: Pair plot between deformation properties and fracture properties on the complete dataset.

test specimen is 10 mm × 8 mm × 0.1 mm, where the fracture surface has an area of 10 mm×8 mm and the thickness of the deformed region around the fracture plane is assumed to be 0.1 mm [1]. The black solid line indicates the value at which Charpy impact energy would equal deformation energy. It is clearly shown that Charpy impact energy is significantly higher than the deformation energy. Since Charpy impact energy is the sum of the deformation energy and the fracture energy, this indicates that the fracture energy takes the main role in controlling the Charpy impact energy.

In addition, the loading mode differs in tensile test and Charpy test, and strain rate in a Charpy test is significantly higher than in a tensile test [13]. This leads to the difficulty in finding the relations between Charpy impact energy and tensile properties. Nevertheless in the following section, it is investigated with various techniques whether prediction of Charpy impact energy only from tensile properties can be done, followed by the analysis based on the prediction models.

6.4. Prediction of Charpy impact energy

In order to predict the Charpy impact energy based on the four tensile properties, i.e. yield strength, ultimate tensile strength, uniform elongation and area under

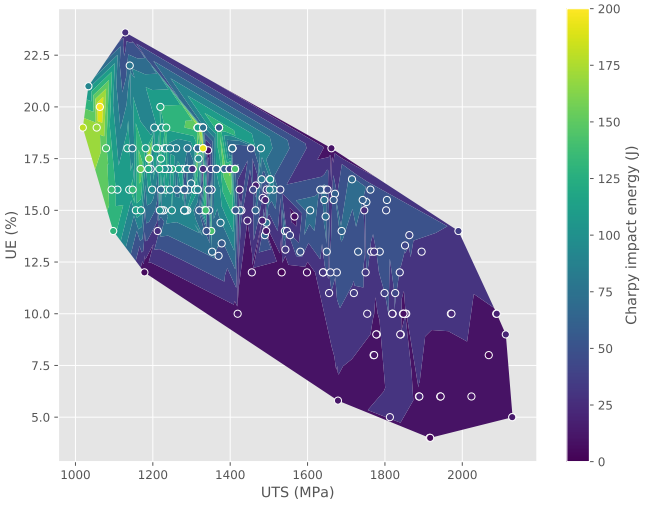


Figure 6.4: Charpy impact energy colored scatter plot between UTS and UE.

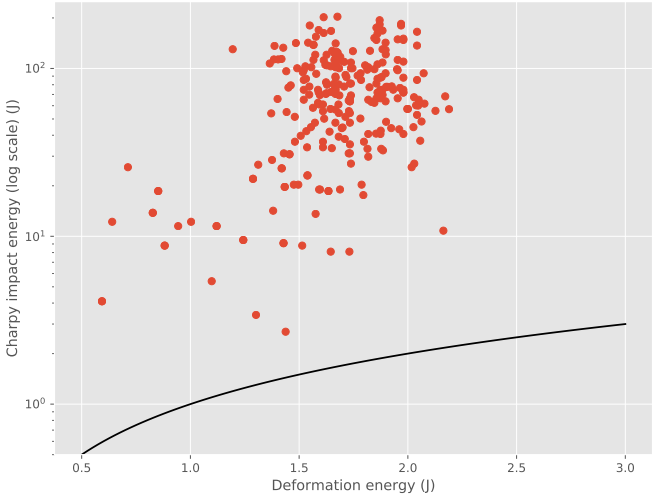


Figure 6.5: Relation between the estimated deformation energy and Charpy impact energy. The solid line represents the location where deformation energy equals to Charpy impact energy.

the tensile curve, the 263 entries of data containing both Charpy impact energy and tensile test properties are divided into 80% training dataset and 20% testing dataset. The machine learning algorithms that were previously applied in the chapters 4&5 (details in Chapter 4 Section 4.B) are applied and tested. Among them random forest gives the best prediction accuracy of the Charpy impact energy based only on the tensile test properties. The *RandomForestRegressor* from *scikit-learn Machine Learning in Python* library [14] is used to apply the random forest regression model. After repeating the random forest model 10 times on different separations of training dataset and testing dataset, the root mean square error (RMSE) is 17 J on the training dataset, and 30 J on the testing dataset. One of the random forest prediction plots of the Charpy impact energy is shown in Fig. 6.6. It shows the experimental Charpy impact energy versus the random forest model predicted values for both the training dataset and the testing dataset. The legend on the bottom right shows a typical value of the experimental uncertainty of the Charpy test ($\pm$20 J) [15], which gives an indication of the best achievable prediction accuracy of the random forest model. The general trend of the points follows the diagonal line that indicates the perfect prediction, while still some of the samples have significant deviations between the real value and the predicted value. Fig. 6.7 shows the error distribution of the random forest prediction on the testing dataset of all ten models, which is calculated as the difference between real testing values and random-forest predicted testing values. Fig. 6.7(a) shows the histogram of error combined with two normal distributions with mean of zero and standard deviation of 20 J and 30 J respectively. It can be seen that the error distribution of the prediction model is slimmer than the normal distribution with standard deviation of 30 J. This indicates that more samples are predicted well when the error is close to zero and meanwhile there are also more large errors comparing to the normal distribution. In Fig. 6.7(b), the barplot shows the histogram of the absolute error while the stepped line plot shows the accumulated distribution. When comparing with the experimental uncertainty of the Charpy test, i.e. $\pm$20 J, nearly 60% of the testing samples have a smaller prediction error. More than 75% of the testing samples have a prediction error lower than 30 J. There is indeed about 1% which have a prediction error larger than 80 J. Therefore, random forest model gives reasonable prediction accuracy of Charpy impact energy from tensile properties when comparing to the accuracy that can be achieved in experiments.

6.5. Interpretation of random forest model

To understand the machine learning model, i.e. to explain the algorithm, there are several methods which can be applied. Here three different methods, partial dependence plot, accumulated local effects plot and the shapley additive explanations, are introduced, where the latter two methods are mainly used to further understand the random forest model. Both the advantages and the disadvantages of these methods are discussed in the following sections.

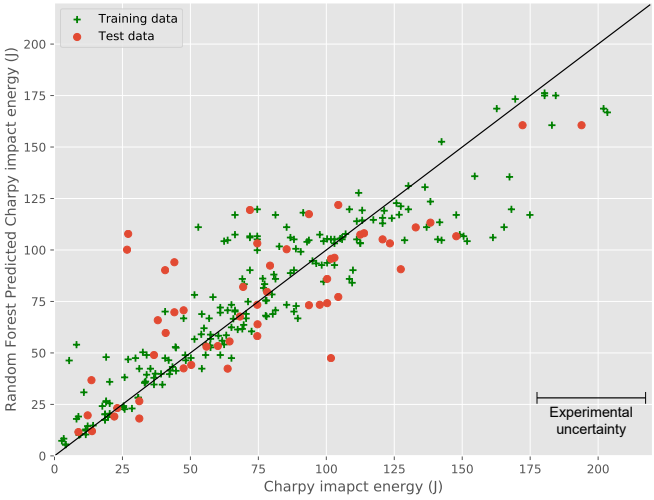
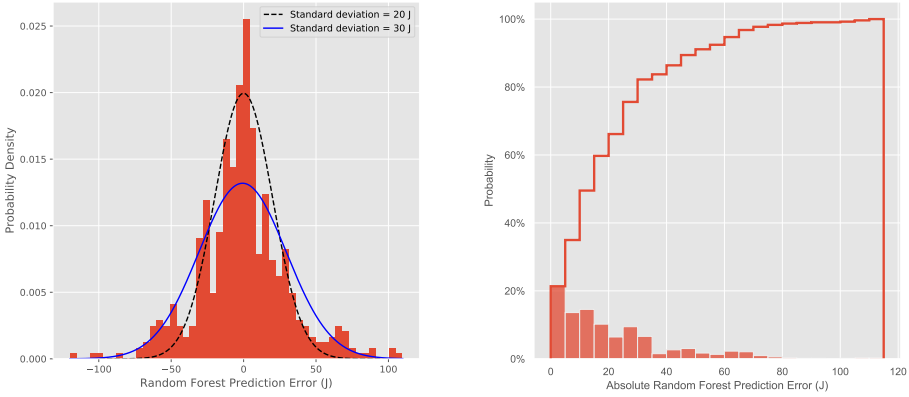


Figure 6.6: Random Forest predicts Charpy impact energy based on tensile test properties.



(a) (b)

Figure 6.7: Error distribution of the Random Forest prediction on the testing dataset of all ten models. Two normal distributions with standard deviation of 20 J and 30 J are shown in (a). (b) shows the absolute error distribution and the accumulated distribution.

6.5.1. Partial Dependence Plot (PDP)

In order to understand the relation between deformation properties and fracture properties, prediction is not enough. The partial dependence plot is introduced first, since it is the most intuitive method to explain the prediction model. In general terms, the partial dependence function for regression is defined as:

$$f_S(x_S) = E\left[f(x_S, X_C)\right] = \int p(x_C) f(x_S, x_C) dx_C. \tag{6.1}$$

f is the machine learning model which we want to explain. The complete feature set, which is used in the machine learning model, in the present case containing σ_y, σ_u, ϵ_u and A_d, is divided into two feature sets S and C. Set S contains the features which we intend to plot in the PDP, that usually only contains one or two features, and set C contains the other features. $p(x_C)$ is the marginal density of X_C. Partial dependence marginalizes the machine learning model output over the other features in set C, in order to give the function which only depends on the features in set S. The partial dependence function f_S is calculated by the Monte Carlo method, i.e. estimating f_S by calculating the averages in the data set as:

$$\hat{f}_S(x_S) = \frac{1}{n} \sum_{i=1}^{n} f(x_S, x_C^{(i)}), \tag{6.2}$$

where n is the number of data points in the dataset and $x_C^{(i)}$ are the actual feature values in the set C for the ith data point. This partial function gives the average machine learning model output depending on the given value/values for feature S that we are interested in [16]. The partial dependence plot can be calculated with the *PDPbox* library [17–19].

Partial dependence has a relatively intuitive computation process. The partial dependence function gives the average prediction output by assuming that all training data points have the same value of each feature in feature set S. Hence the interpretation of the partial dependence plot is easy. It represents the change of the predicted parameter if only the features in feature set S are changing, which leads to the relation of the predicted parameter with the features in feature set S.

This intuitive computation process also leads to its huge disadvantage, if the features in set S have correlations with features in set C. In this case, the UTS, YS, UE and A_d all have correlations to a certain degree. This makes the partial dependence plot unrealistic. For example, when the feature in set S is UTS, the partial dependence function will calculate the average value of all data points with a certain low UTS value. In this case, the data points with high YS values will become meaningless since YS and UTS have a positive correlation. Therefore, another interpretation method in the following section should be considered to overcome this disadvantage.

6.5.2. Accumulated Local Effects (ALE) Plot

The disadvantage of the partial dependence plot which is mentioned in the previous section can be avoided using the accumulated local effect plot [20]. The formula

to define ALE is:

$$f_{x_S,ALE}(x_S) = \int_{\min(x_S)}^{x_S} E\left[\frac{\delta f(X_S, X_C)}{\delta X_S} | X_S = z_S\right] dz_S - C$$
$$= \int_{\min(x_S)}^{x_S} \int p(x_C|z_S) \frac{\delta f(x_S, x_C)}{\delta x_S} dx_C dz_S - C, \tag{6.3}$$

where the constant C is used to center the effect of ALE which will be explained in detail later. $\frac{\delta f(x_S,x_C)}{\delta x_S}$ is the local effect, i.e. the gradient, of f with respect to x_S. Set $S = \{j\}$ in this case only contain one feature, where j is the index of the feature.

While PDP averages the prediction over the marginal distribution of x_C, i.e. without any reference to x_S, ALE plot applies a conditional distribution. In the meantime, ALE does not average the prediction, but averages the changes over the conditional distribution $p(x_C|x_S)$, which means that the distribution of x_C at given restrictions for x_S, i.e. the distribution of x_C with reference to x_S. If there is no correlation between features in x_C and x_S, the conditional distribution is just the marginal distribution. But if there is dependence between the features in x_C and x_S, the conditional distribution is different from the marginal distribution. This helps to isolate the effect of x_S from the influence of other correlated features in x_C. Finally, the last integral reflects the word 'accumulated' in ALE. The local effects are integrated over the range of x_S to reveal the global effect of the feature in set S.

In reality, not all models that we want to explain have a gradient, like the random forest model used in this case. However, the estimation of ALE plots works on models both with and without gradients by dividing the features into several intervals based on the quantiles and calculating the prediction difference between the starting point of the interval and the end point of the interval. The uncentered ALE is calculated by:

$$\hat{\tilde{f}}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in [z_{k-1,j}, z_{k,j})} \left[f(z_{k,j}, x_{\setminus j}^{(i)}) - f(z_{k-1,j}, x_{\setminus j}^{(i)})\right], \tag{6.4}$$

where j is the index of the feature and $k_j(x)$ is the index of the interval in which x is located. $n_j(k)$ is the number of data points within the interval. The interval, in which x is located, has the range $[z_{k-1,j}, z_{k,j})$. The calculation of $f(z_{k,j}, x_{\setminus j}^{(i)}) - f(z_{k-1,j}, x_{\setminus j}^{(i)})$ is the prediction difference between starting point of the interval and the end point of the interval, where $x_{\setminus j}^{(i)}$ indicates the features other than the jth feature. The first sum in Eq. (6.4) accumulates the average prediction difference from the 1st interval to the interval in which x is located. For instance, if x is in the third interval, the uncentered ALE calculates the sum of the prediction difference in the first interval, second interval and third interval. Here for instance, for the feature space x_S containing UTS in Fig. 6.8(a), UTS is divided into 8 intervals (bins) which contain the same number of data points. If a UTS value is within the second interval, 12.5% to 25%, we first calculate the prediction difference when we replace the UTS

value with the grid UTS value at 12.5% and 25% for all data points located within this interval. Then the mean difference is calculated by dividing by the number of data points in this interval. Then due to the first sum on the left, the same calculation has to be performed again in the first interval (0 to 12.5%). Finally the two mean prediction differences are summed to get the uncentered ALE value of UTS at this specific value.

To make the mean of the effect to be zero, uncentered ALE is centered by subtracting the mean value of all uncentered ALE values as:

$$\hat{f}_{j,ALE}(x) = \hat{\hat{f}}_{j,ALE}(x) - \frac{1}{n}\sum_{i=1}^{n} \hat{\hat{f}}_{j,ALE}(x_j^{(i)}), \tag{6.5}$$

where n is the number of data points. To interpret the ALE values, it is recognized as the main effect of the feature at a certain value comparing to the average prediction over the whole dataset.

Fig. 6.8 shows the ALE plot for the random forest model on all four tensile test properties. The plot is calculated with the *ALEPython* library [20, 21]. The bins in Fig. 6.8, indicated by the vertical white lines, represent the intervals in the ALE calculation process. The calculation of ALE averages 50 replicas, each of which randomly samples 60% of the data points from the dataset. From the four ALE plots, it can be seen that UTS and YS have negative effects on Charpy impact energy while UE and A_d have positive effects on Charpy impact energy. Comparing the magnitude of the ALE values of YS and UTS, it shows that UTS has a much more pronounced negative effect than YS. From the ALE plot, it is clear how each of the input feature affects the Charpy impact energy.

6.5.3. SHapley Additive exPlanations (SHAP)

SHAP (SHapley Additive exPlanations) is a method proposed by Lundberg and Lee [22] to explain the individual predictions of a model. SHAP calculates the contribution of each feature on the prediction output for every sample in the dataset. It is based on the Shapley values [23], which came out from a unique question in game theory: How to assign payouts to each player, while players work in coalition to win a certain profit, depending on their contribution to the total payout?

For a function f, which uses p features as the complete feature set F, in order to calculate the SHAP value for feature i, first all subsets M of the set $F \setminus \{i\}$ (the complete feature set without the feature i) are enumerated. The number of possibilities to select a set $M \subseteq F \setminus \{i\}$ is $\binom{p-1}{|M|}$, where the number of features in set $F \setminus \{i\}$ is $p-1$ and the number of features in set M is $|M|$. Then the SHAP value $\phi_{i,x}$ for feature i to an individual data point x is calculated with:

$$\phi_{i,x} = \frac{1}{p}\sum_{M \subseteq F\setminus\{i\}} \frac{1}{\binom{p-1}{|M|}} C(i|M)(x_{M\cup\{i\}}), \tag{6.6}$$

where for each set $M \subseteq F \setminus \{i\}$, $C(i|M)(x_{M\cup\{i\}})$ is the contribution of feature i calculated by $C(i|M)(x_{M\cup\{i\}}) = f(x_{M\cup\{i\}}) - f(x_M)$. It means the prediction difference

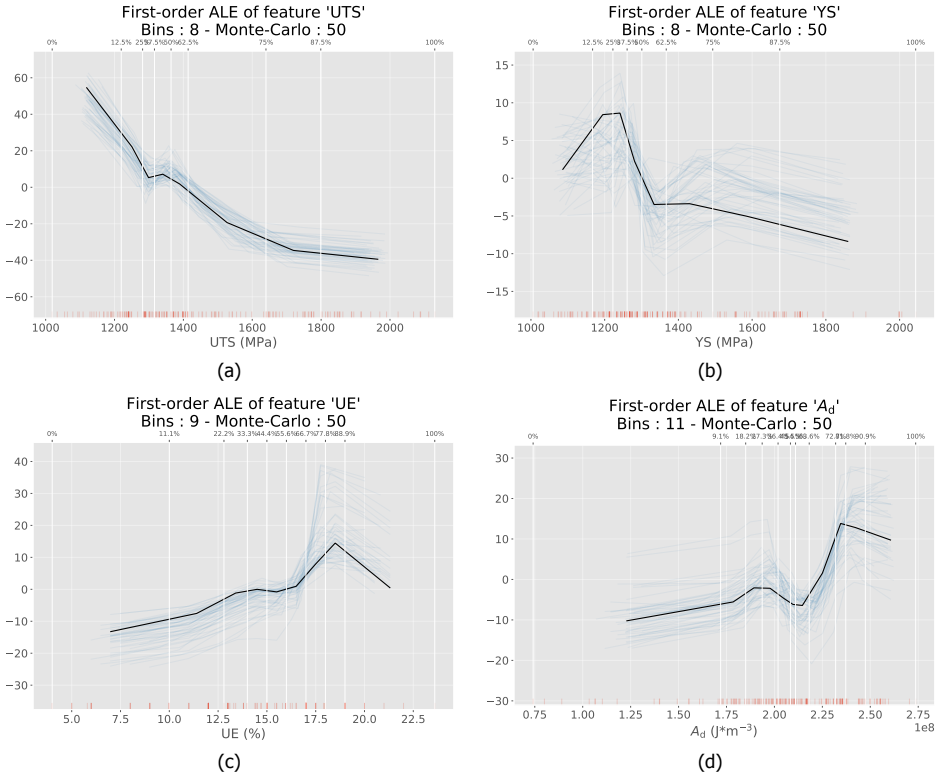


Figure 6.8: Accumulated Local Effects Plot of random forest model for UTS (a), YS (b), UE (c) and A_d (d) on Charpy impact energy.

between before adding and after adding the feature i into the set M of features. The question here is that most prediction models do not accept the arbitrary patterns of missing values, so that $f(x_M)$ needs to be estimated. In order to estimate the SHAP with tree-based models, Lundberg *et al.* [24, 25] proposed *TreeSHAP*. In this case, $f(x_M)$ is approximated with the conditional expectation $E(f(x)|x_M)$, where the features in set M are fixed and the features in the complementary set are sampled from the complete dataset. Hence the prediction difference is calculated as:

$$C(i|M)(x_{M\cup\{i\}}) = E[f(X)|X_{M\cup\{i\}} = x_{M\cup\{i\}}] - E[f(X)|X_M = x_M]. \tag{6.7}$$

Here in the present case, the p features are σ_y, σ_u, ϵ_u and A_d. To calculate the SHAP value for feature σ_u, the set $F \setminus \{i\}$ contains σ_y, ϵ_u and A_d. The subsets of set $F \setminus \{i\}$ can contain one, two or three ($|M|$) elements. Hence the number of subsets of set $F \setminus \{i\}$ of size 1, 2 and 3 are $\binom{3}{1} = 3$, $\binom{3}{2} = 3$ and $\binom{3}{3} = 1$ respectively. For each subset, the difference of prediction before and after adding σ_u into this subset is calculated. Then the mean difference for $|M| = 1$, $|M| = 2$ and $|M| = 3$ is calculated. Finally these three mean differences are summed up and divided by the number of features (four).

According to the name, Additive exPlanations, SHAP values obey the local accuracy property, such that:

$$f(x) = E(f(X)) + \sum_{i=1}^{p} \phi_{i,x}, \tag{6.8}$$

where $E(f(X))$ is the expectation (average prediction) value for all data points. This means that the sum of the SHAP values for each feature plus the baseline (expectation of all model outputs) equals to the final prediction value of the corresponding data point. Hence the prediction of each data point is correctly distributed to each feature.

Fig. 6.9 shows the summary plot of SHAP value for the random forest model. This plot combines the feature importance and the feature effects. The features are ranked from top to bottom with decreasing importance, which is calculated by adding the absolute value of all SHAP values of the corresponding feature over the whole dataset. The color indicates the feature value, from low (blue) to high (red). The points which are overlapping are jittered in the vertical direction to give the sense of the distribution of SHAP values for each feature. The high UTS values contribute negatively to Charpy impact energy, while UE and A_d behave in the opposite direction. This confirms the finding in Section 6.3. The effect of YS is not so clear in this case since there are low-YS (blue) points across the range of SHAP values and the high-YS (red) points accumulate near the zero-SHAP range.

Since UTS is the most important feature according to SHAP importance ranking, the SHAP dependence plot of UTS with the SHAP values is shown in Fig. 6.10, with the two-color interactive features UE and YS. In these two subplots (a) and (b), the scatter plot is the UTS value versus the corresponding SHAP values for UTS, while the color indicates the value of UE in (a) and YS in (b). The scatter

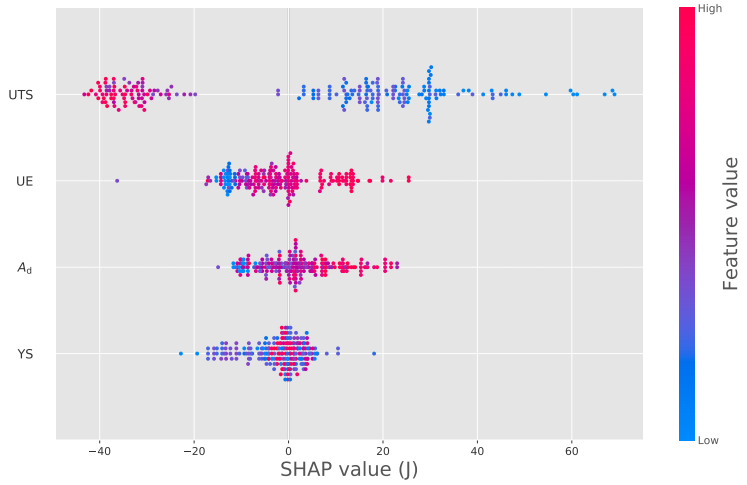


Figure 6.9: Summary of SHAP value magnitudes over all samples for the random forest model prediction of Charpy impact energy.

plots in Fig. 6.10(a) and (b) show that with the increase of UTS until 1400 MPa, the influence of UTS on Charpy impact energy transforms from positive to negative comparing to the average prediction. For UTS higher than 1400 MPa, UTS keeps the same SHAP values, while they are corresponding to the sample with low uniform elongation values (blue points Fig. 6.10(a)) and high YS values (red points Fig. 6.10(b)). This indicates that for samples with low UE and high YS, increasing UTS does not significantly affect the Charpy impact energy.

6.6. Discussion

With the analysis of relations between properties, besides some of the qualitative trends, it is difficult to express the relationship between Charpy impact energy and tensile properties directly by comparing Charpy impact energy with each of the tensile properties. By calculating the deformation energy from the tensile test, it is shown that fracture energy may take the main role in controlling the Charpy impact energy. Meanwhile, due to the difference in loading mode and the significant deviation in stain rate between tensile test and Charpy test, it is even harder to direct relate these properties.

Based on the dataset which is acquired, the prediction of Charpy impact energy based on four tensile test properties is applied with random forest model. The accuracy of predicting the Charpy impact energy is ± 30 J. With the reference of [15], which shows a ± 20 J uncertainty on experimental Charpy impact energy considering all kinds of uncertainty processes during the test, the random forest model shows a prediction, relying on the tensile test properties only, with an uncertainty that is larger than experimental determination, but still displays a clear trend over the considered range of 200 J.

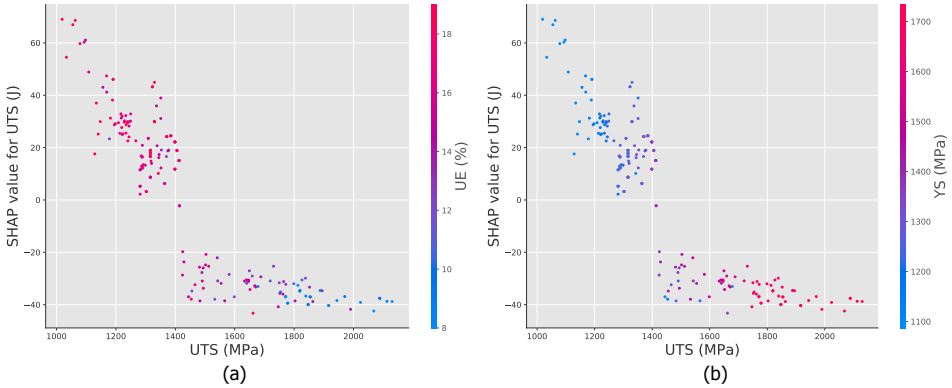


Figure 6.10: SHAP value interactive plot between UTS and UE (a) and YS (b) for the random forest model prediction of Charpy impact energy.

The interpretation of the random forest model indicates that the tensile test properties have different influence on Charpy impact energy, which is in line with the physical background. UTS and YS has profound negative impact on Charpy impact energy, while UE and A_d have significant positive influence. This indicates that the energy absorbed during Charpy test is highly correlated with the overall strength and ductility of the material. The reason why materials with higher UTS and lower UE have smaller Charpy impact energy could be connected to not only the plastic deformation, but also to the fracture behavior during Charpy tests. Materials with high UTS combined with low UE display mostly or purely brittle fracture. This very limited plastic deformation leads to lower Charpy impact energy. From Fig. 6.10, it can been seen that the influence of UTS on Charpy impact energy becomes stable while reaching 1400 MPa. The effect of UTS is still negative at this point, but it does not change if the UTS increases further. These samples are recognized as the ones with high YS and low UE. This shows that in case of the brittle fracture, UTS does not affect the Charpy impact energy.

6.7. Conclusions

This chapter combines deformation properties and fracture properties. With the obtained dataset, detailed data analysis is performed on the relationships between Charpy impact energy and tensile properties. Meanwhile, based on the data of tensile test properties and Charpy impact energy, random forest model is applied to realize the prediction of Charpy impact energy based on ultimate tensile strength, yield strength, uniform elongation and area under the tensile curve. While random forest model gives a reasonable prediction, the interpretation of the model is conducted with the accumulated local effects plot and the shapley additive explanations. The following conclusions are drawn from the study.

- Fracture energy takes a more important role than deformation energy in controlling the Charpy impact energy. Meanwhile, due to the difference in loading

mode and stain rate, it is difficult to find direct relations between Charpy impact energy and tensile properties.

- Random forest model predicting Charpy impact energy based only on UTS, YS, UE and A_{d} gives an accuracy of $\pm 30\,\mathrm{J}$.
- ALE and SHAP can be used to interpret prediction models in the materials science field, which is not commonly adopted. This gives insight into the prediction model.
- UTS and YS have a profound negative effect on Charpy impact energy while UE and A_{d} have significant positive effect on Charpy impact energy.
- For Advanced High Strength Steels that display brittle fracture the effect of UTS on Charpy impact energy is limited.

References

[1] R. O. Ritchie, *The conflicts between strength and toughness,* Nature Materials **10**, 817 (2011).

[2] Y. Kimura, T. Inoue, and K. Tsuzaki, *Tempforming in medium-carbon low-alloy steel,* Journal of Alloys and Compounds **577**, S538 (2013).

[3] Y. Kimura, T. Inoue, F. Yin, and K. Tsuzaki, *Inverse Temperature Dependence of Toughness in an Ultrafine Grain-Structure Steel,* Science **320**, 1057 (2008).

[4] T. Inoue, Y. Kimura, and S. Ochiai, *Shape effect of ultrafine-grained structure on static fracture toughness in low-alloy steel,* Science and Technology of Advanced Materials **13**, 035005 (2012).

[5] Y. Tomita, *Development of fracture toughness of ultrahigh strength, medium carbon, low alloy steels for aerospace applications,* International materials reviews **45**, 27 (2000).

[6] J. Liao, M. Hotta, K. Kaneko, and K. Kondoh, *Enhanced impact toughness of magnesium alloy by grain refinement,* Scripta Materialia **61**, 208 (2009).

[7] W. Cao, M. Zhang, C. Huang, S. Xiao, H. Dong, and Y. Weng, *Ultrahigh Charpy impact toughness (~450J) achieved in high strength ferrite/martensite laminated steels,* Scientific Reports **7**, 41459 (2017).

[8] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, *Machine learning for molecular and materials science,* Nature **559**, 547 (2018).

[9] G. Conduit, *Mechanical properties of some steels, id: 153092 - version 3,* (2017).

[10] S. Kalpakjian and S. Schmid, *Manufacturing Engineering & Technology* (2009) pp. 56–87.

[11] J. Rychlewski, *On hooke's law,* Journal of Applied Mathematics and Mechanics **48**, 303 (1984).

[12] G. Gottstein, *Physical Foundations of Materials Science* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).

[13] E. Lucon, *Experimental assessment of the equivalent strain rate for an instrumented charpy test,* Journal of Research of the National Institute of Standards and Technology **121**, 165 (2016).

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python,* Journal of Machine Learning Research **12**, 2825 (2011).

[15] M. Lont, *The determination of uncertainties in charpy impact testing,* UNCERT COP **6** (2000).

[16] C. Molnar, *Interpretable Machine Learning* (Lulu.com, 2020).

[17] J. Li, *python partial dependence plot toolbox,* (2018).

[18] J. H. Friedman, *Greedy function approximation: a gradient boosting machine,* Annals of statistics , 1189 (2001).

[19] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, *Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation,* Journal of Computational and Graphical Statistics **24**, 44 (2015).

[20] D. W. Apley and J. Zhu, *Visualizing the effects of predictor variables in black box supervised learning models,* (2016).

[21] M. Jumelle, A. Kuhn-Regnier, and S. Rajaratnam, *Python accumulated local effects package,* (2020).

[22] S. M. Lundberg and S.-I. Lee, *A unified approach to interpreting model predictions,* in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017) pp. 4765–4774.

[23] L. Shapley, *A value for n-person game,* Contributions to the theory of games **2**, 307 (1953).

[24] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, *From local explanations to global understanding with explainable AI for trees,* Nature Machine Intelligence **2**, 56 (2020).

[25] S. M. Lundberg, G. G. Erion, and S.-I. Lee, *Consistent individualized feature attribution for tree ensembles,* (2018).

Appendix

6.A. Cleaned experimental dataset

Table 6.A.1: Cleaned dataset from public datasets [9], the area under the tensile curve is calculated from yield strength σ_y, ultimate tensile strength σ_u and uniform elongation ϵ_u by $A_d = \sigma_y * \epsilon_u + 0.5 * (\sigma_u - \sigma_y) * \epsilon_u$.

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
1	1414.5	1465.5	15.0	216.0	64.4
2	1342.9	1661.7	18.0	270.4	10.8
3	1367.0	1493.8	14.4	206.0	72.5
4	1150.6	1156.1	15.0	173.0	113.2
5	1999.5	2068.4	8.0	162.7	3.4
6	1207.1	1228.5	15.0	182.7	79.3
7	1106.5	1134.1	18.0	201.7	111.9
8	1908.5	2023.6	6.0	118.0	11.5
9	1078.3	1138.9	16.0	177.4	113.9
10	1575.7	1714.9	16.5	271.5	68.2
11	1207.8	1223.7	16.0	194.5	101.7
12	1529.6	1639.8	16.0	253.6	67.4
13	1791.4	1915.4	4.0	74.1	4.1
14	1263.6	1279.5	16.0	203.4	104.4
15	1149.3	1223.7	19.0	225.4	104.4
16	1149.9	1188.5	17.0	198.8	71.9
17	1274.7	1333.2	15.0	195.6	138.3
18	1222.3	1351.1	16.0	205.9	167.4
19	1506.8	1541.3	14.0	213.4	78.0
20	1036.9	1054.9	19.0	198.7	169.5
21	1738.0	1841.0	9.0	161.1	22.0
22	1343.6	1405.6	18.0	247.4	88.1
23	1194.0	1223.7	15.0	181.3	76.9
24	1692.9	1764.5	13.0	224.7	36.6
25	1288.4	1316.0	16.0	208.4	70.5
26	1600.5	1675.6	12.0	196.6	47.5
27	1200.2	1276.0	17.0	210.5	100.3
28	1304.3	1466.2	16.2	224.4	17.6
29	1309.0	1370.0	12.8	171.5	54.0
30	1254.0	1316.0	19.0	244.2	66.4
31	1236.1	1249.2	15.0	186.4	100.3
32	1736.3	1839.6	9.0	160.9	22.0
33	1218.8	1260.2	18.0	223.1	85.4
34	1260.9	1288.4	17.0	216.7	47.5
35	1113.4	1190.6	17.5	201.6	202.0
36	1081.7	1147.9	16.0	178.4	132.9

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
37	1197.5	1229.2	18.0	218.4	126.8
38	1329.8	1384.9	17.0	230.7	99.0
39	1378.0	1412.5	17.0	237.2	128.8
40	1555.8	1646.7	16.0	256.2	65.1
41	1267.8	1329.8	18.0	233.8	172.2
42	1285.7	1320.1	15.0	195.4	58.3
43	1226.4	1329.8	17.0	217.3	27.1
44	1437.3	1459.3	16.0	231.7	40.9
45	1660.5	1887.9	6.0	106.5	18.6
46	1329.8	1384.9	17.0	230.7	151.8
47	1456.5	1512.4	16.0	237.5	77.7
48	1143.7	1350.4	17.0	212.0	44.1
49	1615.7	1746.6	15.0	252.2	25.8
50	1731.0	1855.0	10.0	179.3	19.7
51	1309.1	1364.2	17.0	227.2	40.7
52	1138.9	1194.7	17.0	198.4	74.6
53	1212.6	1240.2	17.0	208.5	111.2
54	1404.2	1429.0	15.0	212.5	44.5
55	1254.0	1281.5	15.0	190.2	65.1
56	1183.0	1218.2	15.0	180.1	96.3
57	1512.4	1649.5	13.0	205.5	57.4
58	1213.3	1233.3	16.0	195.7	112.5
59	1593.7	1730.8	13.0	216.1	55.7
60	1272.6	1318.1	17.5	226.7	98.3
61	1274.7	1329.8	19.0	247.4	147.8
62	1371.1	1398.7	17.0	235.4	168.1
63	1475.8	1529.6	16.0	240.4	44.1
64	1267.8	1322.9	19.0	246.1	180.3
65	1340.8	1353.2	13.0	175.1	65.8
66	1390.4	1413.8	15.0	210.3	80.3
67	1451.7	1494.4	16.0	235.7	32.5
68	1091.4	1140.3	22.0	245.5	76.6
69	1316.0	1371.1	19.0	255.3	136.9
70	1267.8	1288.4	17.0	217.3	100.3
71	1309.8	1345.6	15.0	199.2	61.0
72	1260.9	1288.4	17.0	216.7	63.7
73	1715.6	1894.8	13.0	234.7	33.2
74	1247.1	1302.2	17.0	216.7	35.3
75	1593.7	1655.7	11.0	178.7	31.2
76	1357.3	1398.7	17.0	234.3	66.4
77	1343.6	1405.6	18.0	247.4	40.7
78	1729.4	1770.7	8.0	140.0	11.5

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
79	1083.8	1108.6	16.0	175.4	113.2
80	1662.0	1889.0	6.0	106.5	18.6
81	1384.9	1489.6	13.8	198.3	62.4
82	1722.5	1846.5	10.0	178.4	9.1
83	1316.0	1371.1	19.0	255.3	85.4
84	1243.6	1267.1	17.0	213.4	109.5
85	1683.0	1990.0	14.0	257.1	37.1
86	1675.0	1779.0	9.0	155.4	9.5
87	1260.9	1316.0	18.0	231.9	147.8
88	1389.7	1542.7	13.1	192.1	33.9
89	1254.0	1316.0	19.0	244.2	149.1
90	1378.0	1412.5	17.0	237.2	63.7
91	1316.0	1371.1	19.0	255.3	165.4
92	1378.0	1412.5	17.0	237.2	93.6
93	1626.7	1799.0	11.0	188.4	39.7
94	1260.9	1288.4	17.0	216.7	66.4
95	1309.1	1364.2	17.0	227.2	101.7
96	1324.3	1351.8	16.0	214.1	51.5
97	1195.4	1220.9	16.0	193.3	69.8
98	1233.3	1281.5	16.0	201.2	36.6
99	1406.9	1481.4	16.5	238.3	77.7
100	1061.7	1079.0	18.0	192.7	142.4
101	1005.9	1033.5	21.0	214.1	91.5
102	1167.2	1220.9	17.5	209.0	125.8
103	1383.5	1424.9	15.0	210.6	69.1
104	1311.0	1375.0	14.4	193.4	78.0
105	1368.4	1424.2	16.0	223.4	20.3
106	1313.0	1378.0	13.4	180.3	55.0
107	1371.1	1398.7	17.0	235.4	74.6
108	1874.1	1970.5	10.0	192.2	23.0
109	1393.0	1641.0	12.0	182.0	30.8
110	1035.6	1097.6	14.0	149.3	130.2
111	1618.5	1752.1	15.4	259.5	61.7
112	1420.0	1455.9	12.0	172.6	14.2
113	1130.0	1212.6	14.0	164.0	26.7
114	1309.1	1364.2	17.0	227.2	29.8
115	1579.2	1659.1	12.0	194.3	44.7
116	2006.0	2130.0	5.0	103.4	13.8
117	1260.9	1316.0	18.0	231.9	81.3
118	1729.4	1943.0	6.0	110.2	8.8
119	1343.6	1405.6	18.0	247.4	109.8
120	1274.7	1329.8	19.0	247.4	150.5

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
121	1750.1	1812.8	5.0	89.1	25.8
122	1639.8	1786.6	13.0	222.7	50.3
123	1254.0	1316.0	19.0	244.2	97.6
124	1684.6	1754.2	10.0	171.9	28.5
125	1998.1	2087.7	10.0	204.3	18.6
126	1309.1	1364.2	17.0	227.2	65.1
127	1212.6	1240.2	17.0	208.5	120.7
128	1473.8	1643.3	15.5	241.6	74.7
129	1189.9	1234.0	18.0	218.2	100.3
130	1154.8	1228.5	17.0	202.6	105.1
131	1415.2	1502.7	16.5	240.7	78.2
132	1300.8	1351.1	14.0	185.6	141.7
133	2005.0	2129.0	5.0	103.4	13.8
134	1212.6	1240.2	17.0	208.5	105.8
135	1729.4	1853.4	10.0	179.1	19.7
136	1109.3	1178.2	12.0	137.2	5.4
137	1258.8	1289.1	15.0	191.1	103.0
138	1555.8	1598.5	12.0	189.3	8.8
139	1212.6	1240.2	17.0	208.5	74.6
140	1212.6	1240.2	17.0	208.5	103.0
141	1107.9	1147.2	18.0	203.0	82.7
142	1172.0	1213.3	18.0	214.7	88.8
143	1999.0	2089.0	10.0	204.4	18.6
144	1274.7	1336.7	15.0	195.9	138.3
145	1274.7	1329.8	19.0	247.4	71.9
146	1198.0	1299.0	16.3	203.5	80.0
147	1506.8	1546.8	14.0	213.8	38.0
148	1150.6	1168.5	17.0	197.1	154.6
149	1285.0	1484.1	15.6	216.0	31.2
150	1222.3	1244.3	16.0	197.3	92.9
151	1247.1	1302.2	17.0	216.7	54.2
152	1198.2	1240.2	17.0	207.3	80.7
153	1672.9	1852.0	13.3	234.4	42.6
154	1343.6	1405.6	18.0	247.4	71.9
155	1267.8	1322.9	19.0	246.1	184.4
156	1434.5	1502.0	16.0	234.9	88.9
157	1153.4	1219.5	20.0	237.3	121.3
158	1635.0	1803.1	15.0	257.9	48.3
159	1288.4	1309.1	16.0	207.8	74.6
160	1212.6	1240.2	17.0	208.5	101.7
161	1288.4	1309.1	16.0	207.8	112.5
162	1731.0	1944.0	6.0	110.2	8.8

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
163	1113.4	1190.6	17.5	201.6	162.7
164	1793.0	1917.0	4.0	74.2	4.1
165	1288.4	1316.0	16.0	208.4	56.9
166	1230.6	1262.2	16.0	199.4	69.4
167	1399.4	1479.3	18.0	259.1	93.6
168	1141.7	1203.7	19.0	222.8	93.6
169	1269.0	1483.0	14.5	199.5	19.0
170	1462.7	1533.0	12.0	179.7	2.7
171	1073.5	1093.4	16.0	173.4	136.3
172	1267.8	1285.0	15.0	191.5	86.8
173	1267.8	1316.0	18.0	232.5	161.3
174	1688.1	1750.1	12.0	206.3	33.5
175	1021.0	1128.8	23.6	253.7	27.1
176	1390.4	1413.8	15.0	210.3	54.2
177	1391.8	1639.8	12.0	181.9	30.8
178	1216.8	1249.8	17.0	209.7	124.1
179	1254.0	1286.4	16.0	203.2	71.2
180	1557.1	1646.7	14.7	235.5	69.6
181	1288.4	1316.0	16.0	208.4	47.5
182	1260.9	1316.0	18.0	231.9	174.9
183	1212.6	1240.2	17.0	208.5	141.0
184	1233.3	1281.5	16.0	201.2	33.9
185	1316.0	1371.1	19.0	255.3	52.9
186	1232.0	1566.2	14.7	205.7	8.1
187	1416.6	1554.4	13.8	205.0	42.0
188	1205.8	1281.5	16.0	199.0	19.0
189	1194.0	1235.4	17.0	206.5	127.4
190	1243.0	1289.8	18.0	228.0	77.3
191	1267.8	1316.0	18.0	232.5	105.8
192	1254.0	1281.5	15.0	190.2	97.6
193	1371.1	1398.7	17.0	235.4	130.2
194	1329.8	1384.9	17.0	230.7	62.4
195	1232.0	1490.7	15.5	211.0	19.0
196	1371.1	1398.7	17.0	235.4	108.5
197	1169.2	1181.6	18.0	211.6	111.2
198	1212.6	1244.3	18.0	221.1	90.6
199	2043.6	2112.5	9.0	187.0	20.3
200	1260.9	1288.4	17.0	216.7	81.3
201	1254.0	1316.0	19.0	244.2	112.5
202	1254.0	1281.5	15.0	190.2	74.6
203	1493.1	1632.9	16.0	250.1	57.4
204	1019.0	1019.0	19.0	193.6	180.3

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
205	1172.0	1198.2	16.0	189.6	94.9
206	1267.8	1316.0	18.0	232.5	71.9
207	1267.8	1316.0	18.0	232.5	86.8
208	1553.7	1670.8	15.8	254.7	60.3
209	1212.6	1240.2	17.0	208.5	88.1
210	1260.9	1316.0	18.0	231.9	66.4
211	1645.3	1719.7	11.0	185.1	36.6
212	1188.5	1230.6	17.0	205.6	103.0
213	1167.9	1234.7	19.0	228.2	127.4
214	1267.8	1288.4	17.0	217.3	108.5
215	1243.0	1292.6	16.0	202.8	70.5
216	1288.4	1309.1	16.0	207.8	123.4
217	1233.3	1281.5	16.0	201.2	55.6
218	1267.8	1288.4	17.0	217.3	69.1
219	1081.1	1678.6	5.8	80.0	12.2
220	1288.4	1309.1	16.0	207.8	103.0
221	1233.3	1281.5	16.0	201.2	61.0
222	1717.0	1772.0	13.0	226.8	33.2
223	1320.8	1493.1	14.0	197.0	13.6
224	1074.9	1342.9	17.9	216.4	8.1
225	1357.3	1398.7	17.0	234.3	88.1
226	1658.4	1762.5	16.0	273.7	57.2
227	1513.7	1651.5	16.0	253.2	44.7
228	1688.1	1863.1	13.8	245.0	43.4
229	1357.3	1398.7	17.0	234.3	40.7
230	1674.3	1777.6	9.0	155.3	9.5
231	1515.8	1606.7	15.0	234.2	74.6
232	1126.5	1188.5	17.0	196.8	120.7
233	1267.8	1329.8	18.0	233.8	183.0
234	1378.0	1412.5	17.0	237.2	142.4
235	1724.0	1848.0	10.0	178.6	9.1
236	1584.7	1688.1	14.0	229.1	63.2
237	1731.0	1820.0	10.0	177.6	25.4
238	1729.4	1819.0	10.0	177.4	25.4
239	1288.4	1316.0	16.0	208.4	89.5
240	1451.7	1503.4	16.5	243.8	99.0
241	1875.0	1972.0	10.0	192.4	23.1
242	1304.3	1338.7	14.0	185.0	51.5
243	1553.7	1667.4	15.5	249.6	57.4
244	1247.1	1302.2	17.0	216.7	63.7
245	1282.2	1344.2	16.0	210.1	61.0
246	1657.7	1826.5	11.0	191.6	42.3

Table 6.A.1: (continued)

No.	YS (MPa)	UTS (MPa)	UE (%)	A_d (MJ*m^{-3})	K_V (J)
247	1666.0	1838.3	12.0	210.3	39.2
248	1174.7	1216.8	17.0	203.3	120.7
249	1032.1	1063.1	20.0	209.5	203.4
250	1254.0	1281.5	15.0	190.2	85.4
251	1267.8	1288.4	17.0	217.3	81.3
252	1559.9	1742.5	15.5	255.9	60.1
253	1731.0	1772.0	8.0	140.1	11.5
254	1267.8	1329.8	18.0	233.8	193.9
255	1099.6	1444.5	14.5	184.4	20.3
256	1625.4	1805.2	15.5	265.9	55.9
257	1137.5	1208.5	18.0	211.1	117.3
258	1247.1	1302.2	17.0	216.7	31.2
259	1274.7	1329.8	19.0	247.4	74.6
260	1088.6	1419.3	10.0	125.4	12.2
261	1105.8	1169.2	15.0	170.6	107.1
262	1524.8	1581.3	13.0	201.9	50.2
263	1185.1	1453.8	18.0	237.5	48.1

7

Conclusions, general discussion and recommendations

This thesis is based on the research of the project "MICtoMEC: Extensive quantification of microstructure features and statistical relations with mechanical behaviour -from statistical relations to physical understanding-", which was carried out under project number S41.5.14547a in the framework of the Partnership Program of the Materials innovation institute M2i (www.m2i.nl) and the Technology Foundation TTW, which is part of the Netherlands Organization for Scientific Research (www.nwo.nl). This project aims at the development of relations between the intricate and 3D features of multi-phase metallic microstructures of Advanced High-Strength Steels (AHSS) and the mechanical properties of the material, as well as a physical interpretation of these relations.

Due to the complexity of the relevant aspects of the microstructures and various mechanical properties, this thesis focuses on a triangular relation as shown in Fig. 7.1, reflecting the comprehensive approach. Hence the thesis is divided into three parts:

- Microstructure to Deformation: The first part of the thesis (Chapter 3) aims at the relationship between microstructure features and a deformation-related property, i.e. hardness.
- Microstructure to Fracture: The second part of the thesis (Chapters 4 and 5) focuses on the microstructure features' influence on fracture-related mechanical properties, i.e. hole expansion capacity and bendability.
- Deformation to Fracture: The third part of the thesis (Chapter 6) connects deformation properties from tensile tests to the fracture property, i.e. Charpy impact energy.

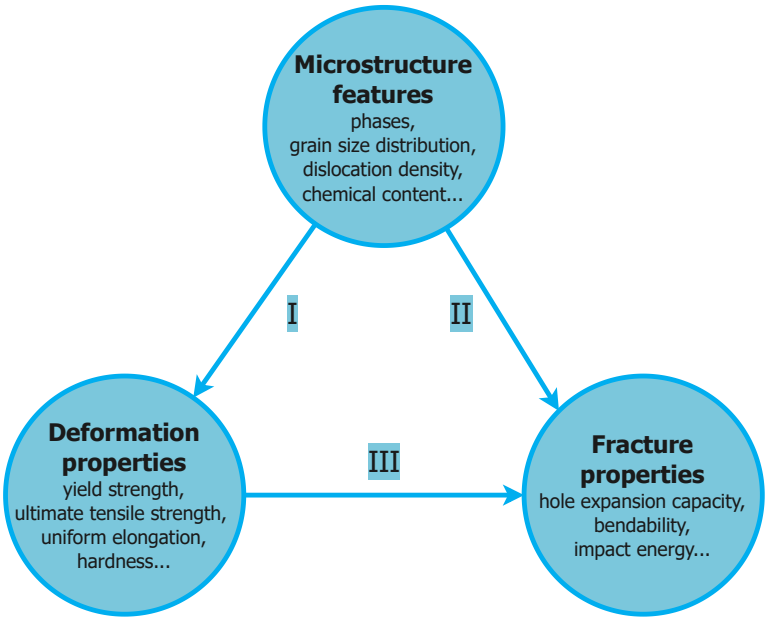


Figure 7.1: Research focus network of this thesis. The three circles represent microstructure features, deformation properties and fracture properties, respectively. The three arrows I, II and III represent the three parts of this thesis, each representing the relationship between the corresponding corners.

In all three parts, relationships are investigated based on experimental data, which is not only the experimental dataset obtained with experiments within this project, but also the research datasets gathered from literature and online databases. In order to study the hidden information in those datasets, not only the traditional data analysis methods are used, but also machine learning and the interpretation of the machine learning models are implemented in this thesis. In the following sections, after summarizing the conclusions with general discussion on these three parts, the limitations and recommendations for future research are discussed.

7.1. Conclusions and general discussion

Part I: Microstructure to Deformation

One of the aims of the original project is to further extend the quantitative characterization of the microstructures. The first part of the thesis (Chapter 3) connects the deformation property hardness with grain size, grain size distribution and dislocation density, focusing on the grain size distribution, which is far less discussed than the average grain size, mentioned with the famous Hall-Petch relation (Eq. (3.1)). In order to understand how the combination of grain size distribution and dislocation density influences the hardness of IF steel, a series of IF steel plates were given dif-

ferent microstructures through various heat treatment routes in combination with cold rolling. It is shown that cold rolling plays a more significant role than heat treatment in increasing hardness, due to the decrease of mean grain size and increase of dislocation density. The combined contribution of dislocation density and grain size distribution on hardness of IF steel plates can be expressed by extended version of Hall-Petch relation and Taylor equation, which adds the contribution of a grain size distribution parameter, kurtosis.

In the relation between microstructure features and deformation properties, the key point is the movement of dislocations. In order to have higher hardness, the movement of dislocations has to be limited. Cold rolling increases the dislocation density directly. Smaller grain size leads to higher grain boundary to dislocation ratio. These all contribute to the difficulty for dislocations to move, hence increasing the hardness. Meanwhile, for kurtosis, the positive effect of kurtosis can also be explained with grain boundary density. With increasing kurtosis the tails of the grain size distribution are stronger, which means there are more extremely small or large grains. This leads to higher grain boundary density than for distributions with smaller kurtosis, hence higher grain boundary to dislocation density ratio.

In this part, besides the traditional linear regression methods, LASSO is introduced, which acts as a variable selection tool for the linear model. LASSO adds a shrinkage penalty term to the loss function of the linear regression model and selects the variables based on the importance by changing the penalty term. LASSO is an interesting tool for materials scientists since linear models are widely implemented.

Part II: Microstructure to Fracture

The second part of the thesis (Chapters 4 and 5) focuses on the relation between microstructure features and fracture properties, namely hole expansion capacity and bendability. This part focuses on data acquired from literature to investigate the relation of phase volume fractions and chemical compositions with hole expansion capacity and bendability. In this part, different kinds of machine learning algorithms are implemented to predict fracture properties based on microstructure features. For both hole expansion capacity and bendability, statistics show that not only the volume fraction of certain phases, but also the combinations of certain phases contribute to the fracture properties.

On the relations between fracture properties and phase fractions, ferrite and martensite show the most pronounced impact. Meanwhile, the influence of phase fractions on fracture properties cannot be easily described with monotonic functions. It occurs often that certain combination of phases contributes significantly to fracture properties. Taking into account the hardness difference among the phases, for both bendability and hole expansion capacity, the volume fractions of harder phases contribute positively. At the same time, the fracture initiation and the crack growth are both favored by the strength mismatch of phases. These observations indicate that the fracture properties are highly strength-related.

In this part, for the case of both the bendability and the hole expansion capacity, it is shown that prediction of the fracture-related properties based only on the phase

fractions and chemical content is not an easy task. For hole expansion capacity, a deep learning model achieves a prediction accuracy close to the experimental accuracy, while for bendability, the prediction accuracy is not as good as for the case of hole expansion capacity. This could be on account of the lack of data points and also the variation of bendability even with the same/similar combination of phases. Nevertheless, machine learning prediction models show huge advantages. Even though the dataset used in this part of the thesis cannot be recognized as so-called big data, it is meaningful that this amount of data in materials science field already shows nice prediction ability in some cases with the help of different machine learning prediction models. It is believed that significant improvement can be achieved with the enhancement of the dataset, both in the amount and the quality.

Part III: Deformation to Fracture

The third part of the thesis (Chapter 6) relates deformation properties and fracture properties. In this part, not only the relation analysis and prediction performance are included, but also the detailed interpretation of the machine learning prediction model is covered with various techniques. With the obtained dataset, detailed data analysis is performed on the relationships between Charpy impact energy and tensile properties. Meanwhile, based on the data of tensile test properties and Charpy impact energy, a random forest model is applied to realize the prediction of Charpy impact energy based on ultimate tensile strength, yield strength, uniform elongation and area under the tensile curve. While random forest model gives a reasonable prediction, the interpretation of the model is conducted with the accumulated local effects plot and the shapley additive explanations.

For Charpy impact energy, the fracture energy part takes a more important role than the deformation energy part. This is one of the reasons why it is difficult to predict the Charpy impact energy solely based on the tensile properties. Other possible causes can be the differences in strain rate and the loading modes. Despite the difficulties, random forest model gives a prediction accuracy for Charpy impact energy of $\pm 30\,\mathrm{J}$, which is close to the experimental uncertainty.

Even though most of the machine learning models are like black boxes, there are different methods to dig into the boxes. In this part of the thesis, three different methods are introduced and both the advantages and disadvantages of these methods are discussed along with the interpretation of the random forest model. Besides the prediction of Charpy impact energy, those interpretation methods indeed give more insight into the prediction model, hence helping to understand the physical connections of the tensile properties and Charpy impact energy. The interpretation methods show that ultimate tensile strength and yield strength have negative relation, while uniform elongation and the area under the tensile curve have positive relation with Charpy impact energy. Moreover, the influence of ultimate tensile strength becomes negligible when the advanced high strength steels show brittle fracture. These interpretations indicate that Charpy impact energy is strength-related, but not exclusively dependent on strength. Other factors like fracture mode can also contribute to the magnitude of the Charpy impact energy.

7.2. Recommendations

Microstructure & Mechanical properties

Besides theories like the Hall-Petch relation and the Taylor equation, many studies only investigate the qualitative relations between microstructure features and mechanical properties. Within this thesis, while the focus is on the statistical relations between Microstructure & Mechanical properties, only a limited number of microstructure features and mechanical properties are included and discussed. However, in order to further investigate the relations, it should be better to extend the scope of both microstructure features and mechanical properties.

In the case of microstructure features, it is common to include aspects like chemical content, phase volume fraction, grain size, dislocation density, etc. In this thesis, an uncommon feature like grain size distribution is included. There are certainly more of these that could be utilized in the study of microstructure-properties relations. Possible suggestions are texture, grain shape, grain boundary density, inclusion characteristics, etc. With the characterization of the microstructure features, traditional methods mainly have two strategies, either viewing from the 2D section of the material, such as optical microscopy and scanning electron microscopy (SEM), or measured from 3D but only using the average data from certain volume, such as XRD. With the development of 3D characterization techniques, such as 3D-EBSD, atom probe tomography (APT) and 3D Synchrotron Laminography, 3D laboratory-based X-ray diffraction contrast tomography (LabDCT), more and more 3D microstructure features can be recorded systematically. Hence it will be beneficial and effective to study the influence on the properties directly from these 3D microstructure features.

In the case of mechanical properties, it is indeed good to include different kinds of mechanical testing methods, and meanwhile to take different aspects of the testing results into account. But viewing from a broader scope, the properties of materials are not limited to being only mechanical. It is of great importance to study more angles of the properties of materials, such as corrosion resistance, thermal properties, electrical and magnetic properties, etc.

Materials data management

Data is so important for materials science research that more attention should be given to materials data management. In terms of management of the materials data, there are three aspects during the lifecycle of the data, namely creating, storing and sharing, which shall be taken into consideration.

In the stage of creating data, it is important to keep the standards consistent. It is challenging in the field of materials science to record all data systematically. However, this is the starting point, hence the base of good materials data management. Currently, there are standards for various testing methods of materials, but within those standards, there is no information that focuses on the way of recording the data. The data recording standards should at least specify the standard data entries which need to be recorded for specific experiments.

In the stage of storing data, the database structure and the data safety can be

the two main points to be considered. Since materials data comes from so many different directions, it is challenging to decide on the materials database structure. A good idea would be to have a database that uses different key nodes to connect all the possibly related data entries. On the consideration of data safety, there are many detailed aspects that may need consideration. But this part is more common to all different fields which implement their database. For the materials science field, the guarantee of not losing the data is probably more important than the prevention of data theft, since here the focus is on those data which is intended to be freely used by researchers all over the world.

The last stage of sharing data is certainly the key point of the materials data management. Without sharing, the data created and stored cannot be utilized by more researchers. Taking the materials data volume into consideration, it is certainly impossible to have one place to store all materials data. Hence the sharing stage is like a free trading market. Like the real market and the online shopping websites, some digital marketplaces can be organized to share information about the metadata of the databases from different contributors in the world.

Finally, the motivation of contributing to this materials data management has to be mentioned. The ideal world would be where everybody is willing to contribute to a certain aspect of the management process, but the reality is skinny. Regulations made by authorities may be one of the solutions to this motivation problem. Benefits of the contribution may be another direction for solutions.

Machine learning

Machine learning in general is to use the computer algorithms, which could improve automatically by the use of data and without being explicitly programmed to do so, to enhance the accuracy of the prediction. In this thesis, machine learning algorithms are implemented to predict the mechanical properties based on the obtained datasets.

For the future research directions about machine learning enhancing materials science studies, there are possible directions from the part of the datasets, the machine learning algorithms and the interpretation of algorithms.

From the dataset part, it is almost always beneficial to feed the machine learning algorithms with more training data. However, limited datasets should not be ignored, which could still be valuable if treated properly. The attention should also be drawn to the quality of the datasets, which need proper pre-processing before the training process. This part is closely related to the previous sections on the materials data management, which directly determines the quality of the datasets.

From the part of the machine learning algorithms, it may not be a good idea to create completely new algorithms for materials scientists. However, interdisciplinary collaboration is essential. It is much more efficient to make use of the fast-developing machine learning algorithms from the computer science field or from other fields. There is still a limited number of studies that focus on the implementation of machine learning techniques in the studies of materials science research fields. Besides the prediction from the learning of datasets, machine learning can also accelerate the study of other aspects of materials science. For example, the

use of machine learning algorithms to enhance the quality of the simulation results, or to accelerate the simulation process. Applying machine learning to the design of new materials is also an important topic nowadays.

After the prediction with machine learning algorithms, it is always beneficial to take a step into the algorithms to understand the connections, which may be helpful in the understanding of the hidden physical background. Machine learning algorithms should not remain as black boxes to materials scientist, otherwise, the contribution of machine learning is massively limited in this field. In the last part of this thesis, there are several ways introduced to dig into the algorithms. However, the study is still limited and further enhancement of this part can be of great importance.

Summary

Multi-phase metallic materials such as Advanced High-Strength Steels (AHSS) are of great importance in a wide variety of high-tech industries due to their higher strength compared to conventional (mild) forming steels. The higher strength leads to various advantages in weight, safety and environmental friendliness. In order to develop new AHHS steels, the steel industries make use of multi-scale microstructure modelling to predict mechanical properties from the microstructure features.

This thesis aims at the development of relations between the features of multi-phase metallic microstructures of steels and the mechanical properties of the material. The quantitative characterization of the microstructure will be more involved than is now in use for estimations of the mechanical properties, which is a necessity because of the complexity of multi-phase microstructures. Moreover, the prediction of mechanical properties on the basis of microstructural features will be extended beyond the usual limitation of the yield stress to properties like hole expansion capacity and impact energy. Statistical approaches combined with machine learning algorithms are used to find relations between microstructure features and mechanical properties. Interpretations of the machine learning algorithms are also discussed and the possible deeply embedded relations among mechanical properties are systematically studied.

The research in this thesis deepens the insight into the mechanical behaviour of the microstructure in multi-phase steels and strongly improves property predictions, not only based on microstructure features, but also on deformation properties. Results of this thesis can be directly implemented in microstructure modelling and are directly available for researchers within the steel industry for developing new materials.

As shown in the research focus network in Figs. 1.1 and 7.1, the three main components of this thesis are microstructure features, deformation properties and fracture properties. The three arrows indicate the three parts of this thesis. Hence the thesis is divided into three parts: Microstructure to Deformation, Microstructure to Fracture and Deformation to Fracture.

In Chapter 2, the existing studies about various relationships between microstructure features and mechanical properties are summarized based on the scale of the corresponding microstructure elements, i.e. phases/precipitates, grains and dislocations. Meanwhile, the characteristics of these microstructure elements are also introduced. Both qualitative and quantitative studies are discussed in this chapter. It is shown that most of the studies focus on the qualitative relations between microstructure features and mechanical properties.

Part I contains Chapter 3. Here the influence of microstructure features on the hardness of Interstitial Free steel is studied. Through different heat treatment

routes combined with controlled cold rolling, the grain size distribution and dislocation density are manipulated, which is then qualitatively measured with optical microscopy and X-ray diffraction measurements. With the newly introduced variable selection tool LASSO, it is revealed that dislocation density, mean grain size and kurtosis of grain size distribution are the three microstructure features that most significantly influence the hardness of Interstitial Free steel.

Part II contains Chapters 4 and 5. In Chapter 4, the study focuses on the influence of phase volume fractions and chemical content on the hole expansion capacity. The data analysis is based on collected datasets from the literature. The hole expansion capacity is analyzed in relation to individual phases, combination of phases and number of phases. Meanwhile, several selected machine learning algorithms are implemented to predict the hole expansion capacity with phase volume fractions and chemical content. Among these algorithms, deep learning gives the best prediction accuracy, which is comparable to the accuracy of the experimental determination. Based on the machine learning algorithms, the influence of microstructure features on the hole expansion capacity is also revealed.

Chapter 5 follows a similar structure as the previous chapter, but focuses on the bendability of the material. The dataset for bendability analysis is also collected from the literature. The influence of phase volume fractions on bendability is discussed in detail. Meanwhile, on the prediction of bendability from phase volume fractions and chemical content, two tree-based algorithms give the best result. Although the accuracy of the prediction of bendability is not as good as that of hole expansion capacity, the influence of different microstructure features on bendability is revealed.

Part III contains Chapter 6. Here the focus moves to the relationship between deformation properties and fracture properties. A well-organized dataset containing the Charpy impact energy and the tensile properties is obtained from the online database. The relations between the tensile properties and the Charpy impact energy are studied in detail. Meanwhile, prediction models on the Charpy impact energy based on the tensile properties are implemented. While random forest model gives the best prediction accuracy in this case, the interpretation of the prediction model is extensively discussed in this chapter. With the various interpretation tools, the influence of tensile properties on the Charpy impact energy is analyzed and discussed.

Chapter 7 concludes the whole thesis and gives recommendations for possible future works. The conclusions and general discussion are also following the three parts structure of the thesis. Based on the conclusions obtained from each chapter, the possible physical background and connections are also discussed. Future recommendations are discussed from three directions, i.e. microstructure & mechanical properties, materials data management and machine learning.

Samenvatting

Meerfasige metalen zoals geavanceerde hogesterktestalen (AHSS) zijn van groot belang in een grote verscheidenheid van hightechindustrieën vanwege hun hogere sterkte in vergelijking met conventionele (zachte) vervormingsstalen. De hogere sterkte leidt tot verschillende voordelen op het gebied van gewicht, veiligheid en milieuvriendelijkheid. Om nieuwe AHHS staalsoorten te ontwikkelen, maken de staalindustrieën gebruik van multischaal microstructuurmodellering om de mechanische eigenschappen te voorspellen op basis van de microstructuurkenmerken.

Dit proefschrift beoogt de ontwikkeling van relaties tussen de kenmerken van meerfasige metallische microstructuren van staal en de mechanische eigenschappen van het materiaal. De kwantitatieve karakterisering van de microstructuur betreft een bredere range dan nu gebruikelijk is voor schattingen van de mechanische eigenschappen, hetgeen noodzakelijk is vanwege de complexiteit van meerfasige microstructuren. Bovendien zal de voorspelling van mechanische eigenschappen op basis van microstructurele kenmerken worden uitgebreid tot meer dan de gebruikelijke beperking van de vloeispanning tot eigenschappen zoals Hole Expansion Capacity en kerfslagenergie. Statistische benaderingen gecombineerd met algoritmen voor machine learning worden gebruikt om relaties te vinden tussen microstructuurkenmerken en mechanische eigenschappen. Interpretaties van de machine learning algoritmen worden ook besproken en de mogelijke diep verankerde relaties tussen mechanische eigenschappen worden systematisch bestudeerd.

Het onderzoek in dit proefschrift verdiept het inzicht in het mechanisch gedrag van de microstructuur in meerfasige staalsoorten en verbetert sterk de voorspelling van eigenschappen, niet alleen gebaseerd op microstructuurkenmerken, maar ook op vervormingseigenschappen. Resultaten van dit proefschrift kunnen direct worden toegepast in microstructuurmodellering en zijn direct beschikbaar voor onderzoekers binnen de staalindustrie voor het ontwikkelen van nieuwe materialen.

Zoals te zien is in het onderzoeksschema in Fig. 1.1 en 7.1, zijn de drie belangrijkste onderdelen van dit proefschrift: microstructuurkenmerken, vervormingseigenschappen en breukeigenschappen. De drie pijlen geven de drie onderdelen van dit proefschrift aan. Vandaar dat het proefschrift in drie delen is verdeeld: Microstructuur tot Vervorming, Microstructuur tot Breuk en Vervorming tot Breuk.

In hoofdstuk 2 worden de bestaande studies over verschillende relaties tussen microstructuurkenmerken en mechanische eigenschappen samengevat op basis van de schaal van de relevante microstructuurelementen, d.w.z. fasen/precipitaten, korrels en dislocaties. Ook worden de karakteristieken van deze microstructuurelementen geïntroduceerd. Zowel kwalitatieve als kwantitatieve studies worden in dit hoofdstuk besproken. Het blijkt dat de meeste studies zich richten op de kwalitatieve relaties tussen microstructuurkenmerken en mechanische eigenschappen.

Deel I omvat hoofdstuk 3. Hier wordt de invloed van microstructuurkenmerken op de hardheid van Interstitial Free staal bestudeerd. Door middel van verschillende warmtebehandelingen in combinatie met gecontroleerd koudwalsen worden de korrelgrootteverdeling en de dislocatiedichtheid gemanipuleerd, die vervolgens kwalitatief worden gemeten met optische microscopie en röntgendiffractie. Met het nieuw geïntroduceerde variabele selectiehulpmiddel LASSO, wordt onthuld dat de dislocatiedichtheid, de gemiddelde korrelgrootte en de kurtosis van de korrelgrootteverdeling de drie microstructuurkenmerken zijn die de hardheid van interstitieel vrij staal het meest significant beïnvloeden.

Deel II omvat de hoofdstukken 4 en 5. In hoofdstuk 4 richt de studie zich op de invloed van fasevolumefracties en chemische samenstelling op Hole Expansion. De gegevensanalyse is gebaseerd op verzamelde datasets uit de literatuur. De Hole Expansion Capacity wordt geanalyseerd in relatie tot individuele fasen, combinatie van fasen en aantal fasen. Ook worden verschillende algoritmen voor machine learning geïmplementeerd om de Hole Expansion Capacity te voorspellen op basis van volumefracties van fasen en chemische samenstelling. Van deze algoritmen geeft deep learning de beste voorspellingsnauwkeurigheid, die vergelijkbaar is met de nauwkeurigheid van de experimentele bepaling. Op basis van de algoritmen voor machine learning wordt ook de invloed van microstructuurkenmerken op de Hole Expansion Capacity onthuld.

Hoofdstuk 5 volgt een vergelijkbare structuur als het vorige hoofdstuk, maar richt zich op de buigbaarheid van het materiaal. De dataset voor de buigbaarheidsanalyse is ook verzameld uit de literatuur. De invloed van fasevolumefracties op de buigbaarheid wordt in detail besproken. Voor de voorspelling van buigbaarheid uit fasevolumefracties en chemische samenstelling, geven twee random-forest-gebaseerde algoritmen het beste resultaat. Hoewel de nauwkeurigheid van de voorspelling van de buigbaarheid niet zo goed is als die van de Hole Expansion Capacity, wordt de invloed van verschillende microstructuurkenmerken op de buigbaarheid gepresenteerd.

Deel III omvat hoofdstuk 6. Hier wordt de aandacht gericht op de relatie tussen vervormingseigenschappen en breukeigenschappen. Een goed geordende dataset met de Charpy-slagenergie en de trekeigenschappen wordt verkregen uit een online database. De relaties tussen de trekeigenschappen en de Charpy-slagenergie worden in detail bestudeerd. Daarnaast worden voorspellingsmodellen voor de Charpy-slagenergie op basis van de trekeigenschappen geïmplementeerd. Daar het random forest model in dit geval de beste voorspellingsnauwkeurigheid geeft, wordt de interpretatie van het voorspellingsmodel in dit hoofdstuk uitgebreid besproken. Met de verschillende interpretatie-instrumenten wordt de invloed van de trekeigenschappen op de Charpy-slagenergie geanalyseerd en bediscussieerd.

Hoofdstuk 7 besluit het gehele proefschrift en geeft de aanbevelingen voor mogelijke toekomstige studies. De conclusies en algemene discussie volgen ook de driedelige structuur. Op basis van de conclusies uit elk hoofdstuk worden ook de mogelijke fysische achtergronden en verbanden besproken. Toekomstige aanbevelingen worden besproken vanuit drie richtingen, te weten microstructuur & mechanische eigenschappen, databeheer en machine learning.

Acknowledgements

I certainly did not expect, could not imagine, what a journal and adventure it would be when I packed from Aachen to Delft. It was a great journey to study from Materials Science to Statistics. It was an amazing adventure to experience and discover all that happened during these years. It would not be possible for me to complete this journey and adventure alone. There are all of you who offered me endless help and kindness along the path.

I want to first express my gratitude to my two brilliant promotors Jilt and Geurt. All of these would not even start without you. Thank you for letting me study from you and work with you. I really appreciate all the discussions and meetings we have had. You offered me full of your professional guidance in Materials Science and Statistics. Dank je wel! Geurt, thank you for guiding me into Statistics and for your unbounded knowledge. You helped me a lot in combining machine learning with materials data research. Jilt, thank you for also being my daily supervisor. I appreciate all the time and energy that you devoted to me and my project. You have taught me so much, which is not only about the knowledge, but also the way to discover, the way to think and the way to conquer.

I want to thank M2i and Tata for offering the funding for the MICtoMEC project. Thanks to Viktoria, Lima and Maria for your kind help and support. Thanks to Piet for your insightful perspective from the industry.

Thanks to my committee members for evaluating my thesis, for your fruitful comments and valuable discussions. Thanks Leo, Annoesjka, Dorte, Ulrich and Piet.

I want to thank all my colleagues from the Materials Science and Engineering department. Thank you Prisca for always being there offering your endless support and kind help. Thanks to Maria, Erik, Jan, Yongxiang, Leo and Roumen for the nice discussions and valuable meetings. Thanks to Sander, Richard, Ton, Elise, Ruud, Nico, Hans and Kees for your help with all kinds of experimental work. I also want to express my gratitude to Arthur, Jesus, Hussein, Vitoria, Jhon, Jithin, Jaji and Vibhor for the interesting discussions during the lunchtime and coffee breaks.

A special thanks to my PhD twin, Martina!

There is a group of lovely people without whom my adventure in delft would not be as colorful. Thanks to Sudhi, Chrysa, Alfonso, Konstantina, Javier, Viviam, Carola and Javier, Vangelis and Artemis, Tim and Marilia, Zaloa, Constantinos and Behnam. It was an amazing experience to enjoy such a variety of cultures with you. Thanks for all the parties and craziness we had in delft. During the past years, there is an interesting wechat group with people coming and going. That is my Chinese family here in the Netherlands. Thanks to Yageng, Zhiyuan, Guoping, Meng, Xiaolin, Yuliu, Zhaoying, Jianing, Xiaohui, Quanxin, Gaojie, Jia, Ziyu, Qiang, Chenteng, Yang, Junjie, Chen, Jun and Huihong. It was such a coincidence that

somehow we gathered here in delft and I appreciate all the help you gave me and all the delicious meals we have had together. Meanwhile, I want to thank Yan for being a friend for more than ten years. Thanks to my friends from Aachen: Xiaoxiao, Xiaolong and Yajiao, Liang and Chunhua, Fuhui and Jing.

After finishing my PhD, I joined ARCNL, the wonderful group of Nanoscale Imaging and Metrology. I would like to thank Lyuba for your understanding and support. Thanks to Arie, Ksenia, Ben, Gerwin, Zhouping, Aleksandra, Christos, Perry, Manashee and Tamar. You all help me a lot with my new adventure.

Finally yet importantly, I want to thank my parents, my sister and my brother-in-law. Thanks for your endless love and support even though I am far away. Thanks to my parents-in-law for your unconditional trust and understanding.

Luckiness is mine to meet you, Lu! Thank you for always being with me. Thank you for your patience and indispensable support during my research career. Thank you for your accompany, trust, understanding, sacrifice and love.

谢谢!

Amsterdam, May 2022

Wei

Curriculum Vitæ

Wei Li

Professional Experience

12.2020–present	Advanced Research Center for Nanolithography (ARCNL) Postdoctoral researcher *Project:* Machine learning in metrology and computational imaging *Supervisor:* Dr. L. Amitonova
07.2016–07.2020	Delft University of Technology (TU Delft) Ph.D. Researcher *Project:* A data-driven and machine-learning study on microstructure-property relations in steel *Promotor:* Prof. dr. ir. J. Sietsma Prof. dr. ir. G. Jongbloed
05.2015–05.2016	Max-Planck-Institut für Eisenforschung (MPIE) Research assistant *Project:* Microstructure design and mechanical properties of near-alpha Ti-alloys *Supervisor:* Univ.-Prof. Dr.-Ing. W. Bleck Prof. Dr.-Ing. Habil. D. Raabe Dr.-Ing. S. Sandlöbes

Education

07.2016–07.2020	Ph.D. in Materials Science and Engineering Delft University of Technology, Delft, the Netherlands
10.2013–05.2016	Master of Science in Metallurgical Engineering RWTH Aachen University, Aachen, Germany
09.2009–06.2013	Bachelor of Science in Metallurgical Engineering University of Science and Technology Beijing, Beijing, China

List of Publications

9. **W. Li**, K. Abrashitova, G. Osnabrugge, L. Amitonova, *Generative adversarial network for super-resolution imaging through a fiber*, arXiv preprint, arXiv:2201.00601 (2022).

8. **W. Li**, M. Vittorietti, G. Jongbloed, J. Sietsma, *Microstructure–property relation and machine learning prediction of hole expansion capacity of high-strength steels*, Journal of Materials Science **56**, 19228 (2021).

7. **W. Li**, M. Vittorietti, G. Jongbloed, J. Sietsma, *The combined influence of grain size distribution and dislocation density on hardness of interstitial free steel*, Journal of Materials Science & Technology **45**, 35 (2019).

6. M. Vittorietti, J. Hidalgo, J. Sietsma, **W. Li**, G. Jongbloed, *Isotonic regression for metallic microstructure data: estimation and testing under order restrictions*, Journal of Applied Statistics, 1 (2021).

5. Y. Li, **W. Li**, F.S.L. Bobbert, K. Lietaert, J-H. Dong, M.A. Leeflang, J. Zhou, A.A. Zadpoor, *Corrosion fatigue behavior of additively manufactured biodegradable porous zinc*, Acta Biomaterialia **106**, 439 (2020).

4. M. Vittorietti, P. Kok, J. Sietsma, **W. Li**, G. Jongbloed, *General framework for testing Poisson-Voronoi assumption for real microstructures*, Applied Stochastic Models in Business and Industry **36**, 604 (2020).

3. Y. Li, P. Pavanram, J. Zhou, K. Lietaert, P. Taheri, **W. Li**, H. San, M.A. Leeflang, J.M.C. Mol, H. Jahr, A.A. Zadpoor, *Additively manufactured biodegradable porous zinc*, Acta Biomaterialia **101**, 609 (2019).

2. Y. Li, H. Jahr, X.Y. Zhang, M.A. Leeflang, **W. Li**, B. Pouran, F.D. Tichelaar, H. Weinans, J. Zhou, A.A. Zadpoor, *Biodegradation-affected fatigue behavior of additively manufactured porous magnesium*, Additive Manufacturing **28**, 299 (2019).

1. Y. Li, K. Lietaert, **W. Li**, X.Y. Zhang, M.A. Leeflang, J. Zhou, A.A. Zadpoor, *Corrosion fatigue behavior of additively manufactured biodegradable porous iron*, Corrosion Science **156**, 106 (2019).