

---

Institute of Mechanics and Control Engineering - Mechatronics

Measurement and Control Engineering - Mechatronics

Prof. Dr.-Ing. Oliver Nelles

---

Master Thesis

# Machine Learning Based Fatigue Life Prediction of Metallic Materials

Author: Chirag Angadi  
Mat.-No.: 1725283

Supervising tutor: Prof. Dr.-Ing. Oliver Nelles  
M.Sc. Fabian Schneider

---

University of Siegen  
Faculty of Natural Sciences and Engineering  
Department mechanical engineering  
Paul-Bonatz-Straße 9-11  
D-57068 Siegen

---



**Imprint**

Prof. Dr.-Ing. Oliver Nelles  
Measurement and Control Engineering - Mechatronics  
Department of mechanical engineering  
University of Siegen  
57068 Siegen



# Preface

Metals are widely used in different industries due to properties like strength, durability, malleability, ductility, conductivity, and machinability. These properties make metals indispensable across industries such as construction, automotive, aerospace, electronics, energy, and healthcare. Metal component design plays a critical role in modern engineering, influencing the performance, efficiency, and sustainability of various industrial applications.

Steel is predominantly used in the production of various mechanical components due to its abundant availability, low cost, and high strength. Consequently, significant research has been conducted on steel to enhance its properties and applications. However, other metals, such as aluminum, are utilized in specific applications owing to their high strength-to-weight ratio, excellent corrosion resistance, and superior conductivity.

Metal components are subjected to repeated loading during daily operations, which can eventually lead to the fracture of components over a certain period. Therefore, fatigue fracture analysis of metal components is crucial to prevent major accidents. This thesis primarily focuses on the study of aluminum alloys, building upon the existing body of research conducted on steel.

This thesis marks the culmination of my master's studies in Mechatronics at the University of Siegen. The research journey has been both

challenging and rewarding, offering invaluable insights into data-driven modeling for the fatigue analysis of metals.

The inspiration for this thesis stemmed from the research paper **Analysis of data-driven models for predicting fatigue strength of steel components with uncertainty quantification** [8]. This paper highlights the machine learning models developed for predicting the fatigue life of steel components, demonstrating significant advantages over traditional methods of fatigue analysis. Through extensive exploration and dedication, I have gained a deeper understanding of the subject, which is reflected in this thesis titled **Machine learning based fatigue life prediction of metallic materials**. Building upon the previous research conducted on steel, this thesis presents various methodologies for developing machine learning models aimed at the fatigue fracture analysis of aluminum alloys. I hope that this work will contribute meaningfully to the ongoing research in data-driven modeling for the fatigue analysis of metals. This thesis was completed in February 2025 at Robert Bosch GmbH, Renningen, in collaboration with the University of Siegen.

I would like to express my heartfelt gratitude to my supervisors at the University of Siegen, **Prof. Dr.-Ing. Oliver Nelles** and **M.Sc. Fabian Schneider**, for their unwavering support and guidance. I am equally grateful to my supervisors at Bosch, **Dr. Christian Frie** and **Mr. Dominic Olveda**, as well as my colleagues, for their continuous encouragement and invaluable input throughout this thesis.

I would also like to extend my sincere thanks to my family and friends for their constant emotional support during this journey.

Chirag Angadi  
University of Siegen

# Contents

<b>Preface</b>	<b>III</b>
<b>Glossary</b>	<b>VII</b>
<b>List of Figures</b>	<b>IX</b>
<b>List of Tables</b>	<b>XIII</b>
<b>Abbreviations</b>	<b>XV</b>
<b>1 Introduction and Motivation</b>	<b>1</b>
<b>2 Fundamentals</b>	<b>5</b>
2.1 Fatigue . . . . .	5
2.1.1 Fatigue damage mechanism . . . . .	6
2.1.2 Experimental method and SN curve . . . . .	7
2.1.3 Factors influencing fatigue . . . . .	9
2.1.4 FKM analytical method . . . . .	16
2.2 Machine learning . . . . .	19
<b>3 Database</b>	<b>27</b>
3.1 Data sources . . . . .	27
3.2 Data extraction from data book . . . . .	28
3.3 Features and labels . . . . .	30
3.4 Data analysis . . . . .	31

<b>4 Approaches</b>	<b>37</b>
4.1 Data split and testing methods . . . . .	37
4.2 Methodologies . . . . .	44
4.3 Results . . . . .	53
4.3.1 DoE 1 . . . . .	54
4.3.2 DoE 2 . . . . .	60
4.3.3 DoE 3 . . . . .	61
4.3.4 DoE 4 . . . . .	65
4.3.5 DoE 5 . . . . .	69
4.4 Impact of data volume on model performance . . . . .	71
<b>5 Discussion and Conclusion</b>	<b>75</b>
5.1 Optimal ML models . . . . .	76
5.2 Optimal feature set . . . . .	78
5.3 Optimal training method . . . . .	79
5.4 Future scope . . . . .	80
<b>A Results of other fatigue parameters</b>	<b>83</b>
<b>B Hyper parameters of machine learning models</b>	<b>87</b>
<b>C Bibliography</b>	<b>89</b>

# Glossary

Symbol	Explanation
$G_0$	Stress gradient at hot spot
$K$	Slope of line in finite region of Wöhler curve
$K_t$	Stress concentration factor
$K_f$	Fatigue notch factor
$K_{wk}$	Design factor
$K_{AK}$	Mean stress factor
$L_{type}$	Type of load applied
$N_e$	Number of cycles at knee point in Wöhler curve
$R_m$	Tensile strength (MPa)
$R_{ratio}$	Stress ratio of minimum and maximum stress in cyclic load
$S_d$	Fatigue limit (MPa)
$S_{d, FKM}$	Fatigue limit calculated using FKM analytical method
$T_s$	Scatter of Wöhler Curve in Infinite region



# List of Figures

1.1	Comparison of fatigue analysis methods . . . . .	2
2.1	Different phases of fatigue . . . . .	6
2.2	Laboratory fatigue testing [1] . . . . .	7
2.3	S-N curve obtained from JSMS aluminum data . . . . .	8
2.4	Type of loading . . . . .	10
2.5	Stress strain curve [3] . . . . .	11
2.6	Stress concentration factor . . . . .	12
2.7	Stress gradient [13] . . . . .	13
2.8	Cyclic load . . . . .	15
2.9	Stress cycles with different R-ratios . . . . .	15
2.10	FKM analytical method . . . . .	17
2.11	ML workflow . . . . .	19
2.12	Random data split . . . . .	23
3.1	Steel - Pair plot of $R_m$ , $K_t$ and $S_d$ . . . . .	32
3.2	Steel - Heat map of $R_m$ , $K_t$ and $S_d$ . . . . .	32
3.3	Aluminum - Comparison of $R_m$ , $K_t$ and $S_d$ . . . . .	34
3.4	Aluminum - Heat map of $R_m$ , $K_t$ and $S_d$ . . . . .	34
3.5	Features and target correlation heat map . . . . .	36
4.1	Quartile based random split . . . . .	38
4.2	K-fold testing method . . . . .	40
4.3	Data split method comparison . . . . .	41
(a)	Random forest . . . . .	41

(b) Neural network . . . . .	41
4.4 Conventional and Unconventional K-fold evaluation . . . . .	42
4.5 Different training strategies . . . . .	45
4.6 Separate models for fatigue parameters . . . . .	49
4.7 Early stopping method . . . . .	52
4.8 DoE 1 - Neural Network (Aluminum) . . . . .	54
4.9 DoE 1 - Probabilistic Neural Network (Aluminum) . . . . .	55
4.10 DoE 1 - Error distribution plot (Aluminum) . . . . .	56
(a) Deterministic models . . . . .	56
(b) Probabilistic models . . . . .	56
4.11 DoE 1 - Random Forest (Steel) . . . . .	58
4.12 DoE 1 - Probabilistic Neural Network (Steel) . . . . .	59
4.13 DoE 3 - Prediction plots (Aluminum) . . . . .	63
(a) Neural Network . . . . .	63
(b) Probabilistic Neural Network . . . . .	63
4.14 DoE 3 - Prediction plots with extended data (Steel) . . . . .	64
(a) Linear Regression . . . . .	64
(b) Random Forest . . . . .	64
4.15 DoE 4 - Prediction plots (Steel + Aluminum) . . . . .	67
(a) Neural Network . . . . .	67
(b) Probabilistic Neural Network . . . . .	67
4.16 DoE 4 - Prediction plots (Aluminum) . . . . .	68
(a) Random Forest . . . . .	68
(b) Probabilistic Neural Network . . . . .	68
4.17 DoE 5 - Prediction plots for transfer learning (Aluminum)	70
(a) Neural Network . . . . .	70
(b) Probabilistic Neural Network . . . . .	70
4.18 Performance metrics based on number of data points . . . . .	71
4.19 Fatigue limit ( $S_d$ ) absolute error in feature space $R_m, K_t$	73
4.20 Data count in feature space $R_m, K_t$ . . . . .	73
5.1 DoE comparison for aluminum models . . . . .	77
(a) Deterministic models . . . . .	77

(b) Probabilistic models . . . . .	77
A.1 Slope in finite region ( $K$ ) prediction plots (Aluminum) . . .	84
(a) Random Forest . . . . .	84
(b) Random Forest CI . . . . .	84
A.2 Knee point ( $N_e$ ) prediction plots (Aluminum) . . . . .	85
(a) Random Forest . . . . .	85
(b) Random Forest CI . . . . .	85
A.3 Scatter in infinite region ( $T_s$ ) prediction plots (Aluminum)	86
(a) Random Forest . . . . .	86
(b) Random Forest CI . . . . .	86



# List of Tables

3.1	Data count from different sources . . . . .	28
3.2	Feature description . . . . .	30
3.3	Target label description . . . . .	31
4.1	Seed Sensitivity . . . . .	39
4.2	Aluminum results – Comparison of Hyperparameter-Tuned and Non-Tuned Models . . . . .	43
4.3	Feature sets . . . . .	46
4.4	Design of experiments . . . . .	47
4.5	DoE 1 - Performance metrics (Aluminum) . . . . .	57
4.6	DoE 1 - Performance metrics (Steel) . . . . .	58
4.7	DoE 2 - Performance metrics (Aluminum) . . . . .	60
4.8	DoE 2 - Performance metrics (Steel) . . . . .	60
4.9	DoE 3 - Performance metrics (Aluminum) . . . . .	62
4.10	DoE 3 - Performance metrics (Steel) . . . . .	62
4.11	DoE 3 - Performance metrics with extended data (Steel) .	62
4.12	DoE 4 - Performance metrics (Steel + Aluminum) . . . . .	66
4.13	DoE 4 - Performance metrics (Aluminum) . . . . .	66
4.14	DoE 4 - Performance metrics (Steel) . . . . .	66
4.15	DoE 5 - Performance metrics (Aluminum) . . . . .	69
A.1	Aluminum fatigue parameters performance metrics . . . . .	83
B.1	Hyperparameters - Random Forest. . . . .	87
B.2	Hyperparameters - K Nearest Neighbor . . . . .	87
B.3	Hyperparameters - Support Vector Regressor . . . . .	87

B.4	Hyperparameters - Neural Networks . . . . .	88
B.5	Hyperparameters - Gaussian Process Regressor . . . . .	88

---

# Abbreviations

Abbreviation	Explanation
DoE	Design of experiments
FEA	Finite Element Analysis
FKM	Forschungskuratorium Maschinenbau (Research association for the mechanical engineering industry)
FKM Guideline	Analytical strength assessment of components
GPR	Gaussian Process Regressor
KNN	K Nearest Neighbor
LR	Linear Regression
MAE	Mean Absolute Error
ML	Machine Learning
MSD	Mean Standard Deviation
NLL	Negative Log Likelihood
NN	Neural Network
PNN	Probabilistic Neural Network
RF	Random Forest
RFCI	Random Forest with Confidence Interval
SVR	Support Vector Regressor



# 1 Introduction and Motivation

Aluminum, the third most abundant element in the Earth's crust, is a lightweight, silvery-white metal known for its remarkable properties and wide range of applications. With a density of approximately  $2.7 \text{ g/cm}^3$ , about one third that of steel, it is highly valued in industries where weight reduction is essential, including aerospace, automotive, and construction [16]. Additionally, aluminum is non-toxic, non-magnetic, and naturally resistant to corrosion due to the formation of a protective oxide layer on its surface.

Aluminum's high strength-to-weight ratio further enhances its suitability for structural and load-bearing uses. Additionally, its superior thermal and electrical conductivity makes it an efficient and economical choice for wiring, heat exchangers, and cooling systems. Aluminum's ability to be recycled indefinitely without losing its original properties further solidifies its position as an environmentally sustainable material [16]. These attributes collectively highlight aluminum's indispensable role across a wide range of industries, showcasing its versatility and critical importance in modern engineering and manufacturing.

The design of metallic components plays a crucial role in modern engineering, as it directly influences the performance, reliability, and efficiency of various systems. Effective metal component design involves selecting the appropriate material based on its mechanical properties, such as strength, ductility, fatigue resistance, and corrosion resistance, to ensure it meets the specific demands of the application. Factors such as

geometry, loading conditions, and manufacturing processes are carefully considered to optimize the component's structural integrity and minimize material waste. Innovations in material science and data-driven modeling have further enhanced the ability to design components with improved durability and performance, ensuring their suitability for demanding industrial applications.

In industrial applications, engineering metal components are subjected to continuous cyclic loading during daily operations. This repeated loading can lead to fatigue failure after a certain number of cycles. Since this phenomenon occurs over time, it often goes unnoticed, making it difficult to prevent damage promptly. Fatigue fracture is a significant cause of metal failures, making it a critical consideration in the early stages of component design.

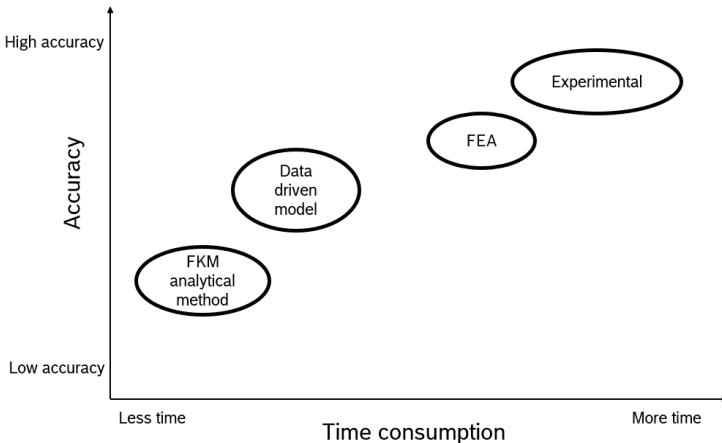


Figure 1.1: Comparison of fatigue analysis methods

Fatigue testing methods play a vital role in evaluating the durability of materials and components under cyclic loading conditions. As shown in Figure 1.1, laboratory experimental methods offer direct measurement of fatigue parameters with high accuracy under specific conditions, but are often time-intensive and costly. Finite Element Analysis (FEA) provides

a simulation-based approach to predict fatigue parameters, allowing efficient and cost-effective assessments for complex geometries. However, this is also time-consuming process.

The FKM analytical approach (see Chapter 2.1.4) provides a standardized analytical framework for fatigue evaluation by integrating empirical data with theoretical models. Although it is an effective and widely used approach, it tends to be overly conservative and less adaptable to unconventional materials. Additionally, this method simplifies complex nonlinear relations into simplex equations for ease of calculation, which results in reduced accuracy compared to other more advanced techniques.

Data-driven modeling that uses machine learning and data analytics offers powerful predictive capabilities while being less time consuming with good accuracy. These models are especially beneficial in the early stages of component design, streamlining the planning of physical tests by optimizing design experiments. By reducing the number of physical tests required, data-driven approaches significantly save both time and costs in evaluation of fatigue behavior.

The significant advantages of the data-driven approach are that it effectively bridges the benefits of other methods. Much of the research in data-driven fatigue analysis has focused on steel, as highlighted in publications such as [4, 5, 8, 10], due to the extensive availability of data for this material. These successful studies on steel pave the way for extending similar research to other metals commonly used in industrial applications, offering opportunities to enhance fatigue analysis across a broader range of materials.

## Research objective

This thesis focuses on aluminum, which has substantial data available, positioning it as a suitable material for further investigation. The primary objective is to identify the most effective data-driven method for developing a machine learning model to predict the fatigue parameters of aluminum. The specific goals of the research are as follows:

- To evaluate and identify the most effective machine learning models, both deterministic and probabilistic, that offer higher accuracy in predicting fatigue parameters for aluminum.
- To expand the existing steel dataset to enhance data variability and quality, ensuring a more thorough and robust analysis.
- To address the data quality challenges of aluminum due to the limited available data.

## Thesis outline

The structure of this thesis is organized as follows: Chapter 1 presents the research, outlining the motivation and objectives of the study. Chapter 2 offers a brief overview of the fundamental concepts of fatigue analysis and machine learning, which form the foundation of this thesis. Given that multiple datasets from various sources will be utilized, Chapter 3 focuses on describing the data sources, the extracted datasets, feature and target definitions, and their correlation. Chapter 4 explores the different approaches employed to achieve the main objectives of the thesis, including three distinct training methods, the design of experiments (DoE) conducted, and a comparison of the results obtained from each DoE. Finally, Chapter 5 summarizes the work carried out in the thesis and presents the conclusions drawn from the analysis and findings of the methodologies used.

## 2 Fundamentals

### 2.1 Fatigue

The occurrence of fatigue fracture is a complex, multi-scale phenomenon influenced by various factors, including the component's manufacturing process, geometric characteristics, applied loads, testing conditions, and environmental factors [21].

#### **What is fatigue limit?**

The fatigue limit is defined as the maximum amplitude of cyclic stress that can be applied to a component without causing fatigue failure, considering  $10^7$  cycles as the threshold for the material's fatigue life. As stated in the book [6], in general practice, components that can withstand  $10^7$  cycles are considered to have an infinite fatigue life.

Section 2.1.1 describes the fatigue damage mechanism that occurs in a component under cyclic load. Furthermore, Section 2.1.2 explains the procedure for obtaining SN curve through experimental methods. Section 2.1.3 discusses the key factors that influence fatigue parameters. Finally, Section 2.1.4 presents the FKM analytical method for determining the fatigue limit.

### 2.1.1 Fatigue damage mechanism

#### Phases of Fatigue

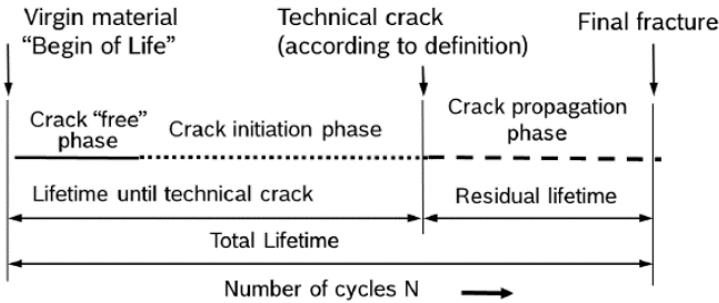


Figure 2.1: Different phases of fatigue

Fatigue fracture is a cumulative process that can be divided into three distinct phases: (1) crack-free, (2) crack initiation, and (3) crack propagation, as illustrated in Figure 2.1. Under cyclic loading, irreversible plastic deformation can occur at high-stress sites, leading to permanent damage in the specimen and the development of cracks. Initially, the specimen remains in the crack-free phase for a certain number of cycles. As the number of loading cycles increases, crack nucleation occurs, marking the onset of the crack initiation phase. Cracks typically originate at notch in the specimen, such as regions of porosity or discontinuities within the specimen. However, fatigue fracture is predominantly a surface crack phenomenon, with cracks originating at the surface in most cases [21].

As loading cycles continue, the crack grows along the plane of maximum stress and through the grain boundary [21]. This results in the formation of a noticeable crack, which is known as a technical crack. Continued cyclic loading leads to further permanent damage, with the remaining time before rupture termed the residual lifetime. This phase corresponds to the crack propagation phase, where crack growth occurs

at an accelerated rate. Eventually, after a certain number of cycles, the specimen ruptures, marking the appearance of a fatigue fracture.

### 2.1.2 Experimental method and SN curve

A traditional and accurate method to determine the fatigue limit is the laboratory testing method, in which samples are subjected to different amplitudes of cyclic stress. The laboratory tests are conducted as follows: initially, an amplitude less than the specimen's tensile strength is selected based on prior knowledge, and the specimen is tested at that amplitude until it either fractures or survives up to  $10^7$  cycles. If the specimen fractures, the test results, including the applied stress amplitude and the number of cycles until fracture, are recorded. As illustrated in the Figure 2.2, the stress amplitude is then reduced, and the same process is repeated until the stress amplitude reaches a point where no fracture occurs, and the specimen survives for  $10^7$  cycles. This combination of stress amplitude and the number of cycles is referred to as runouts, where the specimen did not fracture.

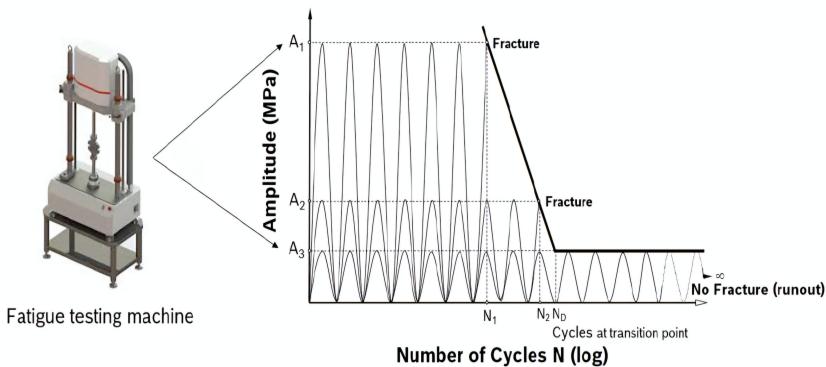


Figure 2.2: Laboratory fatigue testing [1]

As observed in Figure 2.2, the specimen undergoes fatigue fracture at

$N_1$  and  $N_2$  cycles when subjected to stress amplitudes  $A_1$  and  $A_2$ , respectively. However, testing with a stress amplitude of  $A_3$  resulted in no fracture, with the specimen surviving for  $10^7$  cycles. Therefore, in an ideal scenario, as depicted in the figure, stress amplitudes below  $A_3$  MPa cause no damage to the specimen and it is considered as fatigue limit for the tested specimen.

The graph obtained from laboratory experiments Figure 2.2 is known as the S-N curve, which illustrates the relationship between stress amplitudes and the number of cycles to failure. The y-axis of the graph represents the amplitude of cyclic stress applied to the specimen in megapascals, while the x-axis displays the number of applied cycles on a logarithmic scale. However, the S-N curve shown is an idealized case.

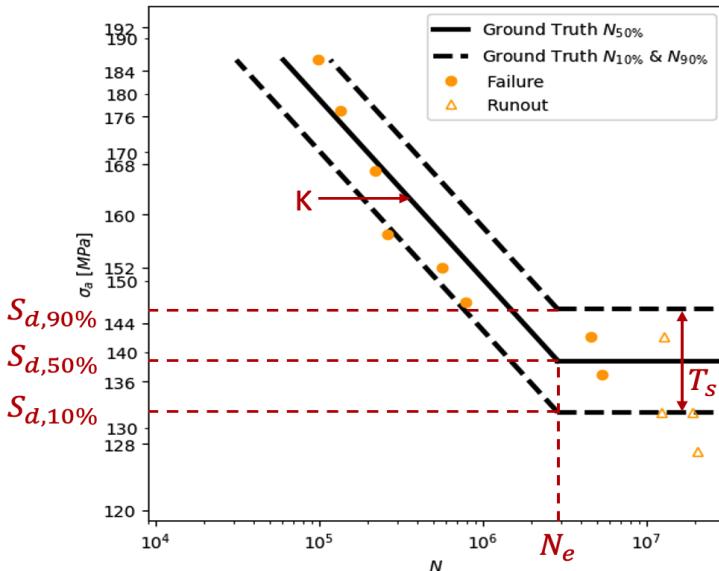


Figure 2.3: S-N curve obtained from JSMS aluminum data

In practical applications, a scatter is observed in the data (Figure 2.3), where specimens tested under the same stress amplitude fail at different

numbers of cycles. This scatter in data is influenced by various factors, including manufacturing processes, testing conditions, and the laboratory environment [20]. Therefore, this variability must be taken into account when evaluating the S-N curve.

### Parameters of S-N curve

The S-N curve (Figure 2.3) obtained from laboratory testing provides crucial fatigue information in the form of four key parameters: the fatigue limit ( $S_d$ ), the knee point ( $N_e$ ), the slope in the finite region ( $K$ ), and the scatter in the infinite region ( $T_s$ ).

The finite region corresponds to the range  $[0, N_e]$ , where fracture data points are observed, depicted as orange dots representing failures in Figure 2.3. Beyond  $N_e$ , the remaining range  $[N_e, \infty]$  is referred to as the infinite region, where no failures occur. These data points are represented by triangles in the figure, indicating runouts.

To account for data scattering, fatigue limits corresponding to 10%, 50%, and 90% failure probabilities are plotted. The fatigue limit  $S_{d,50\%}$  represents the stress level where 50% of the specimens fail. Similarly,  $S_{d,90\%}$  and  $S_{d,10\%}$  indicate the fatigue limits at which 10% and 90% of the specimens fail, respectively, under identical test conditions.

#### 2.1.3 Factors influencing fatigue

Various factors influence the fatigue life of a component, including chemical composition, heat treatment, mechanical properties, component design, and loading conditions. However, this section focuses on a few key factors.

According to previous researches [4, 10, 8] and FKM guideline [11], the fatigue life of a metal component is primarily affected by static tensile

stress (mechanical property), stress concentration factor and stress gradient at the notch (geometric properties), stress amplitude ratio and loading type (test conditions).

### Types of loading ( $L_{type}$ )

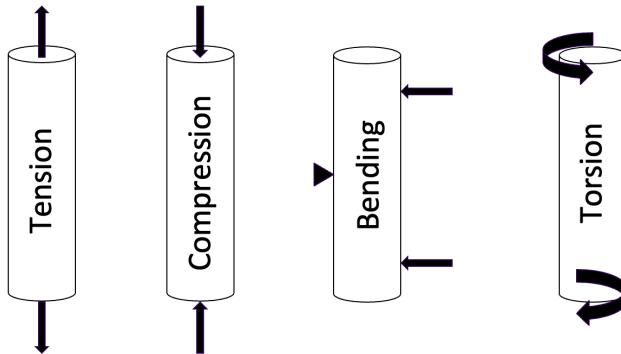


Figure 2.4: Type of loading

Figure 2.4 illustrates that different types of loading have distinct effects on the fatigue limit. Common loading types applied to the specimen include tension-compression, bending, and torsion. Tension-compression and bending result in normal stress (i.e., the force is applied perpendicular to the area resisting it), while torsion generates shear stress, which acts parallel to the area resisting the force. The type of loading applied to the specimen is an important parameter that influences the stress gradient. This parameter indirectly affects the fatigue limit through its impact on the stress gradient.

## Tensile strength ( $R_m$ )

Tensile strength refers to the maximum static stress a specimen can endure before failure. Figure 2.5 depicts the stress-strain behavior under static loading conditions. Initially, strain is directly proportional to stress until the applied stress reaches the yield stress, which defines the elastic zone. Beyond this point, continued increases in stress result in the transition to the plastic zone, where the material experiences both elastic and plastic deformation. This plastic deformation is irreversible and permanent. As the static load further increases, the specimen reaches its ultimate strength, leading to necking, a phenomenon where the net section reduces until failure occurs.

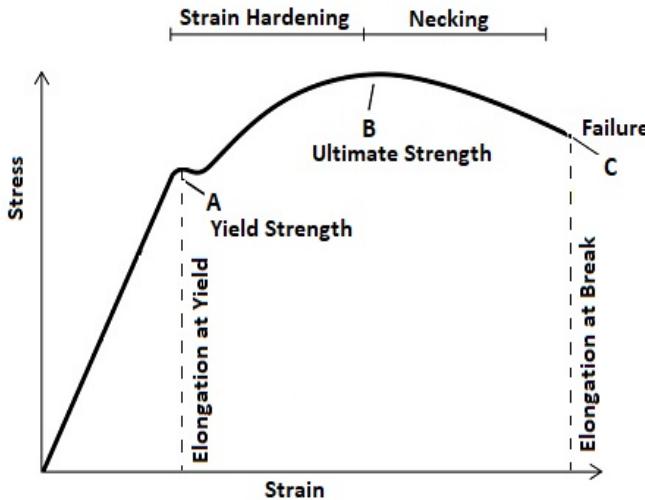


Figure 2.5: Stress strain curve [3]

A research paper [15] presents an approximate relation between the fatigue limit and tensile strength. This relation is derived based on tests conducted on various material alloys:

$$S_d = R_m(C - P \times R_m) \quad (2.1)$$

where  $C$  and  $P$  are constants (refer to [15]).

However, this approximate equation does not apply universally across all regions. Therefore, the **FKM method** employs more accurate analytical equations that take additional parameters into account (see 2.1.4).

### Stress concentration factor ( $K_t$ )

As discussed in the fatigue damage mechanism, the stress concentration factor plays a pivotal role in determining the fatigue limit of a specimen. The failure of the specimen is initiated at its weakest point, typically a geometric discontinuity. The stress concentration factor is defined as the ratio of local stress to nominal stress within the specimen. Local stress refers to the maximum stress occurring at the hot spot of notch at its surface, while nominal stress represents the stress in a specimen without any notches (Figure 2.6).

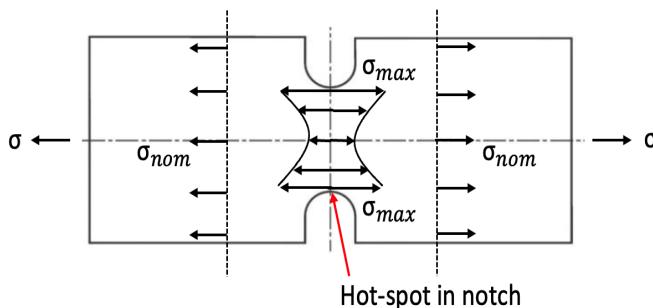


Figure 2.6: Stress concentration factor

$$K_t = \frac{\sigma_{\max}}{\sigma_{\text{nom}}} \quad (2.2)$$

where  $\sigma_{\max}$  is the local stress and  $\sigma_{\text{nom}}$  is the nominal stress.

A higher stress concentration factor **lowers the fatigue limit**, as it increases the likelihood of fracture at notches where the local stress is highest.

## Stress gradient ( $G_0$ )

The stress gradient is another geometric parameter, in addition to the stress concentration factor, that affects the fatigue limit. The stress gradient takes into account the geometry of the notch and the type of load. It primarily influences the fatigue limit by accounting for the sharpness of the notch.

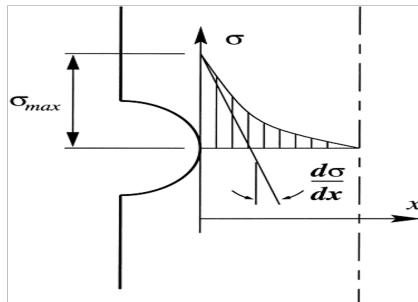


Figure 2.7: Stress gradient [13]

Figure 2.7 depicts the stress distribution in a notched specimen. The maximum stress occurs at the surface of the notch, referred to as the hotspot. As one moves perpendicularly inward from the surface, the stress gradually decreases [13]. The relative stress gradient at the hotspot is denoted as  $G_0$ .

The stress gradient  $G_x$  is defined as the rate of change of stress with respect to the distance  $x$ , given by:

$$G_x = \frac{d\sigma}{dx} \quad (2.3)$$

At the hotspot ( $x = 0$ ), the relative stress gradient  $G_0$  is expressed as:

$$G_0 = \frac{1}{\sigma_{\max}} \frac{d\sigma}{dx} \Big|_{x=0} \quad (2.4)$$

Estimating the stress gradient typically requires Finite Element Analysis (FEA), which involves significant computational effort and time. However, the FKM guideline [11] approximates the relative stress gradient using a simpler equation, as shown below.

$$G_{0,\sigma} = \frac{2}{r}, \quad G_{0,\tau} = \frac{1}{r} \quad (2.5)$$

where  $r$  is the radius of the notch and  $G_{0,\sigma}$  and  $G_{0,\tau}$  are stress gradient for axial and shear stress respectively.

For the remainder of the thesis,  $G_0$  is referred to as the stress gradient, instead of the relative stress gradient, for simplicity.

## Mean stress effect

The specimens are subjected to different amplitudes of cyclic load. The main characteristics are represented by the stress ratio and the mean stress.

Figure 2.8 illustrates the parameters of cyclic loading. The **stress ratio** is defined as the ratio of the minimum stress  $\sigma_u$  to the maximum stress  $\sigma_o$  in a cyclic load. Additionally,  $\sigma_m$  represents the **mean stress**.

$$R_{\text{ratio}} = \frac{\sigma_m - \sigma_a}{\sigma_m + \sigma_a} = \frac{\sigma_u}{\sigma_o} \quad (2.6)$$

The mean stress influences fatigue life depending on the value of  $R_{\text{ratio}}$ . As illustrated in Figure 2.9, when  $R_{\text{ratio}} \geq 0$  and  $R_{\text{ratio}} < 1$ , the specimen undergoes tension-type loading with a mean stress greater than zero. In

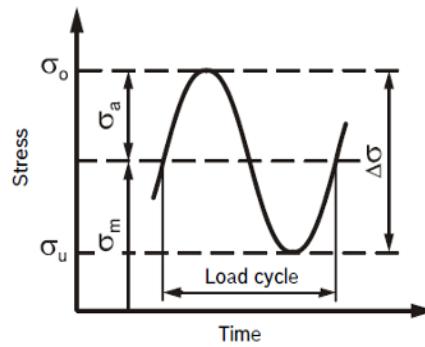


Figure 2.8: Cyclic load

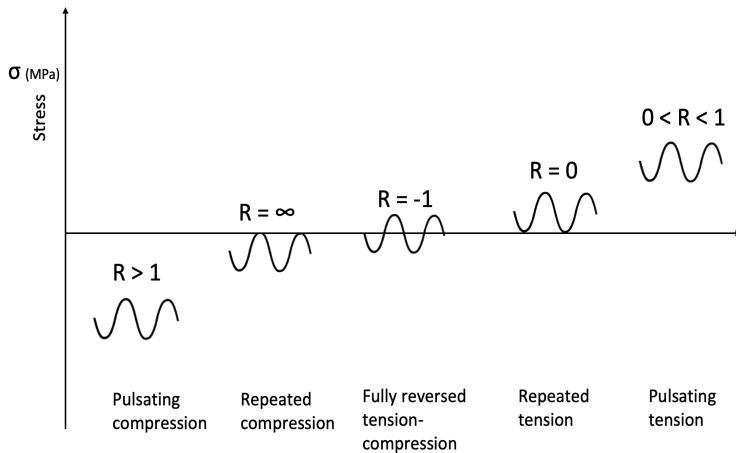


Figure 2.9: Stress cycles with different R-ratios

the case of  $R_{ratio} = -1$ , the specimen experiences fully reversed tension-compression cyclic loading, where the mean stress is zero. For other values of  $R_{ratio}$ , the loading is compressive, resulting in negative mean stress values.

A mean stress greater than zero, associated with tension type load, reduces the fatigue limit. Comparatively, a specimen under fully reversed

tension-compression loading exhibits a higher fatigue limit than one subjected only to tension. However, compressive loading results in the highest fatigue limit.

In addition to these major factors, fatigue parameters are also influenced by chemical composition, heat treatment, and environmental testing conditions. However, the FKM method [11] does not account for these factors, as their contribution to determining fatigue parameters is not significant. While these characteristics are embedded in the tensile strength  $R_m$ , heat treatment also affects the microstructure of the specimen, thereby influencing the fatigue parameters.

#### 2.1.4 FKM analytical method

This section presents the analytical approach outlined in the FKM guideline [11] for determining the fatigue limit. The equations used in this method are derived from the FKM guideline.

This analytical approach considers the influence of four key parameters: tensile strength ( $R_m$ ), stress concentration factor ( $K_t$ ), stress gradient ( $G_0$ ) and stress ratio ( $R_{ratio}$ ) on the fatigue limit.

Figure 2.10 illustrates the fatigue limit calculation process. As discussed earlier, tensile strength plays a significant role in fatigue limit estimation. The fatigue limit is derived by applying various correction factors, such as the design factor ( $K_{wk}$ ), mean stress factor ( $K_{AK}$ ) and fatigue strength factor ( $f_w$ ).

The geometric design factor  $K_{wk}$  is determined using  $K_t$  and  $G_0$ , as shown in the top two blocks of Figure 2.10. The mean stress effect, which depends on  $R_{ratio}$ , is quantified using the factor  $K_{AK}$ , represented in the bottom-left block of the figure. Since the fatigue limit is lower than the tensile strength, a strength factor ( $f_w$ ) is applied to  $R_m$  to determine the fatigue limit for completely reversed stress ( $\sigma_w$ ). Finally,  $\sigma_w$

is corrected using the factors  $K_{wk}$  and  $K_{AK}$  to obtain the final fatigue limit of the specimen.

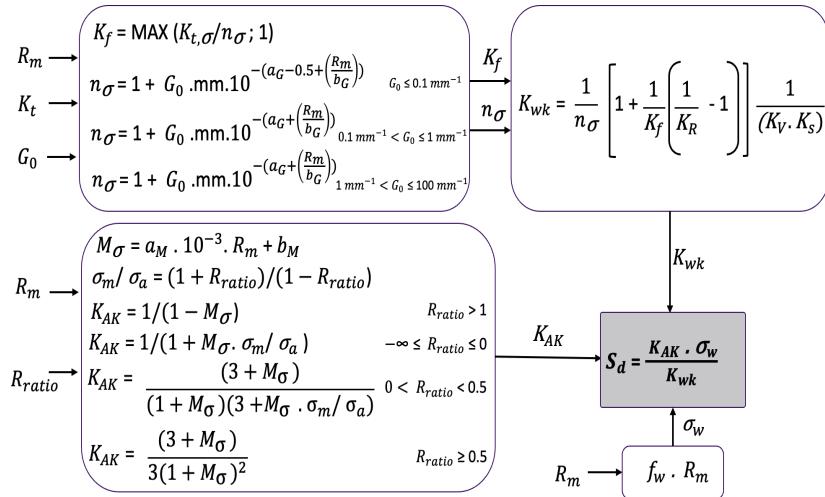


Figure 2.10: FKM analytical method

where,

$K_t$	Stress concentration factor
$K_f$	Fatigue notch factor
$n_\sigma$	$K_t/K_f$ ratio
$K_R$	Roughness factor
$K_v$	Surface treatment factor
$K_s$	Coating factor
$K_{wk}$	Design factor
$M_\sigma$	Mean stress sensitivity
$K_{AK}$	Mean stress factor
$f_w$	Fatigue strength factor for completely reversed stress
$\sigma_w$	Fatigue limit for completely reversed stress
$S_d$	Fatigue limit

$a_G, b_G, a_M, b_M$  Constants based on material

## 2.2 Machine learning

Machine learning (ML) is a powerful decision-making tool that predicts the behavior of unseen data based on historical data. ML models are generally classified into supervised, unsupervised, and semi-supervised learning. In supervised learning, the model is trained using labeled data, meaning it has target labels to learn from and make predictions. Unsupervised learning, on the other hand does not have target labels, the model is trained using all available data to identify patterns and structures. Supervised learning is further divided into regression and classification. Regression involves predicting continuous target values, while classification deals with discrete target labels or distinct classes [14].

Figure 2.11 illustrates the general workflow of automated ML pipeline. This workflow primarily consists of data collection, data preprocessing, model training, and model selection.

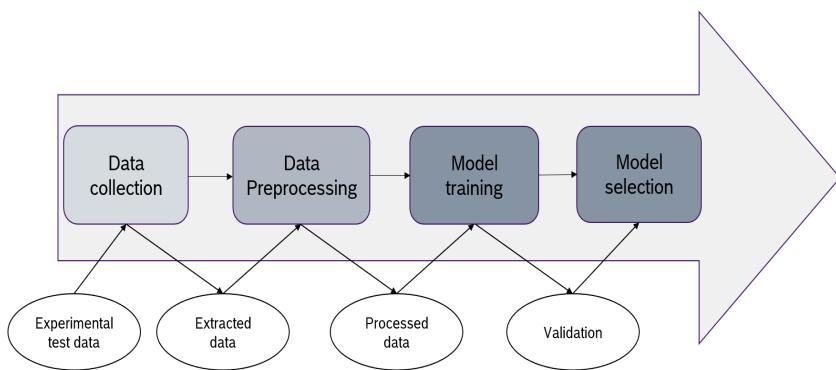


Figure 2.11: ML workflow

### Why is it necessary to provide good data for training?

The first step in the ML workflow is data extraction, where data can come in various forms such as images, tabular data, and signals. Developing a meaningful dataset is crucial for machine learning models because their

accuracy heavily depends on the quality of the data used for training. A well-known saying, "Garbage in, garbage out," highlights this concept—if poor quality data is fed into an ML model, it will produce unreliable and inaccurate predictions.

## Data preprocessing

However, the extracted data is often not in a trainable format, as it may contain unrealistic values or missing data due to human errors, system constraints, or software issues. To make the data suitable for ML models, it must go through a process called data preprocessing in the ML workflow. This involves data cleaning, data transformation, data splitting. Proper data preprocessing is essential for ensuring high-quality input, which leads to better predictions and model reliability [14].

- **Data cleaning:**
  - Instances with missing values in key features and target labels are removed when imputation is not feasible due to limited data availability and high variability.
  - Since machine learning models are sensitive to outliers, they are identified using methods such as the Interquartile Range (IQR) and Z-score. Based on their severity, outliers are either removed or adjusted within a defined range to minimize their impact on model performance.

The IQR method detects outliers based on the spread of the data. It is computed as:

$$IQR = Q_3 - Q_1 \quad (2.7)$$

where  $Q_1$  and  $Q_3$  represent the first and third quartiles, respectively. Any data point falling outside the following bounds is considered an outlier:

$$\text{Lower Bound} = Q_1 - 1.5 \times IQR \quad (2.8)$$

$$\text{Upper Bound} = Q_3 + 1.5 \times IQR \quad (2.9)$$

Similarly, the Z-score method measures how many standard deviations a data point deviates from the mean. The Z-score is calculated as:

$$Z = \frac{x - \mu}{\sigma} \quad (2.10)$$

where  $x$  is the individual data point,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. Data points with  $|Z| > 3$  are considered outliers.

By applying these outlier detection techniques, the dataset is refined to minimize the influence of extreme values, ensuring improved model performance and robustness.

- **Data transformation:**

- Missing values in the dataset are handled using various imputation techniques, such as replacing them with predefined values, mean, median, or the most frequently occurring value.
- Categorical features are encoded using either an ordinal encoder or a one-hot encoder.

- **Why scaling of data is necessary?**

Since different features have varying ranges of values, these

differences in scale can impact how much each feature contributes to predictions. Scaling ensures that all features are treated equally, preventing dominant features from overshadowing others. Additionally, scaling helps in avoiding exploding and vanishing gradient problems in neural networks, leading to more stable training[14].

Generally, two widely used methods of scaling are chosen based on the data distribution:

- \* Normalization: The data is scaled to a range between 0 and 1 using the following formula:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2.11)$$

where  $X$  is the original feature value,  $X_{\min}$  is the minimum value of the feature, and  $X_{\max}$  is the maximum value of the feature.

- \* Standardization: The data is transformed to have a mean of 0 and a standard deviation of 1 using the formula:

$$X_{\text{std}} = \frac{X - \mu}{\sigma} \quad (2.12)$$

where  $\mu$  is the mean of the feature and  $\sigma$  is the standard deviation.

- **Data split:**

- Splitting the dataset into training and testing subsets is essential to evaluate the model's performance on unseen data.
- Various data splitting techniques are used in the machine learning community, including random split, stratified random split, and k-fold split, each with its own advantages and disadvantages [18].

- The random split method is a commonly used data partitioning technique that randomly divides the dataset into training, validation, and testing subsets based on predefined split ratios. The training set is used to learn model parameters, while the validation set helps tune hyperparameters and assess model performance during training. Finally, the model is evaluated on the test set, which remains unseen during training, to measure its generalization ability (Figure 2.12).

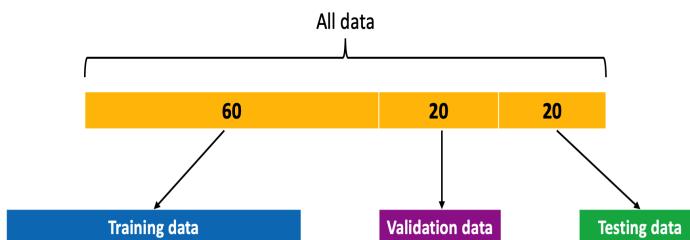


Figure 2.12: Random data split

- Another widely used data splitting method is K-fold cross-validation. Initially, a portion of the data is set aside as test data for the final evaluation of the model. The remaining data is divided into  $k$  folds, where in each iteration, one fold serves as the validation set while the remaining folds are used for training. This process is repeated for all folds, ensuring that each subset is used as a validation set once. The final model parameters are determined by averaging the results across all iterations. This technique is particularly useful when dealing with limited data, as it maximizes data utilization for training and validation (Figure 4.4).

## Machine learning models

After preprocessing, the data is used to train machine learning (ML) models. The choice of ML model depends on factors such as system behavior, use case, and problem complexity. In general, regressor-type ML models can be categorized into deterministic and probabilistic models.

- **Deterministic Models:** These models produce a single, fixed output for a given input. They do not account for uncertainty. Examples include:

- Random Forest
- Neural Networks

- **Probabilistic Models:** These models incorporate uncertainty in their predictions, providing a probability distribution over possible outcomes rather than a single value. They are useful when modeling real-world scenarios with inherent randomness. These models provide mean and variance of prediction.

Examples include:

- Gaussian Process Regression
- Probabilistic Neural Networks

## Evaluation metrics

Machine learning models are evaluated using different metrics depending on whether they are deterministic or probabilistic.

- **Deterministic Models:** These models are typically assessed using error-based and goodness-of-fit metrics.

- **Mean Absolute Error (MAE):** Measures the average absolute difference between actual and predicted values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.13)$$

- **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted values, penalizing larger errors more than MAE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.14)$$

- **R-Squared Score ( $R^2$ ):** Indicates how well the model explains the variance in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.15)$$

where  $\bar{y}$  is the mean of the actual values.

- **Probabilistic Models:** These models are evaluated based on how well they estimate uncertainty and likelihood of observed data.
  - **Negative Log-Likelihood (NLL):** Measures the likelihood of the observed data given the predicted probability distribution. Lower values indicate better predictions.

For a probabilistic model where predictions follow a normal (Gaussian) distribution with mean  $\hat{y}_i$  and standard deviation  $\sigma_i$ , the NLL is given by:

$$NLL = \frac{1}{n} \sum_{i=1}^n \left( \frac{(y_i - \hat{y}_i)^2}{2\sigma_i^2} + \log \sigma_i + \frac{1}{2} \log(2\pi) \right) \quad (2.16)$$

where:

- \*  $y_i$  is the actual observed value
- \*  $\hat{y}_i$  is the predicted mean
- \*  $\sigma_i$  is the predicted standard deviation
- \*  $n$  is the total number of data points

# 3 Database

This chapter outlines the structure and sources of both the existing database and the newly extracted database. Additionally, it provides details about the features and labels used for Machine Learning (ML) models, as well as a comparison of databases. While steel has a wealth of available fatigue test data due to its widespread use in industries, there is also a sufficient amount of published data for aluminum by The Society of Material Science, Japan (JSMS) [2].

## 3.1 Data sources

The data for steel was primarily sourced from the National Institute of Material Science (NIMS)[19] and the Datenbank und Auswertesystem Betriebsfestigkeit (DaBef)[12]. Although this database was utilized for ML model development in previous research[8], it lacks variability as it does not contain extensive data for notched specimens and high tensile strength steel.

In order to enhance the quality of data-driven models for steels, it was essential to incorporate data variability, particularly including data with notched specimens and high tensile strength steels. Therefore, expanding the steel database and obtaining data for aluminum for model development is one of the key objective of this thesis.

The Data book on fatigue strength of metallic materials [2] published by JSMS is the largest available source of data, containing extensive laboratory experimental data for steel with notched specimens. In addition, it includes data for other metals such as aluminum, copper, titanium, and magnesium.

The primary sources of data for this thesis were obtained from:

- Databook on fatigue strength of metallic materials (JSMS) [2]
- National Institute for Material Science (NIMS)[19]
- Datenbank und Auswertesystem Betriebsfestigkeit (DaBef)[12]
- Robert Bosch internal data (RB)

Data source	Steel	Aluminum
JSMS	200	100
NIMS + RB	800	Not available
DaBef	200	30

Table 3.1: Data count from different sources

## 3.2 Data extraction from data book

The data from the Data Book on Fatigue Strength of Metallic Materials (JSMS) [2] was extracted using the Microsoft Azure text reader tool. Each set of experimental data published in the book was converted into a structured tabular format, with clearly defined features and target variables, during the data cleaning process.

The experimental data from JSMS [2] is categorized into several sections, including material properties, chemical composition, heat treatment, tensile test, impact test, fatigue specimen characteristics, and fatigue test results (SN curve).

- The **materials section** provides details about the manufacturer, geometry, and inclusions.
- The **chemical composition section** outlines the mass percentages of various elements present in the specimen.
- The **heat treatment section** describes process parameters such as temperature, cooling environment, cooling duration, grain size, and post-treatment hardness.
- The **tensile test section** includes information on yield stress, tensile stress, elongation, and reduction of area.
- The **impact test section** specifies the testing method and absorbed energy.
- The **fatigue specimen section** details the geometry of fatigue test specimens, notch type, stress concentration factors, specimen surface characteristics, and roughness.
- The **fatigue test section** covers testing conditions, including loading type, machine type, applied stress conditions (constant stress ratio or constant mean stress), loading frequency, environmental conditions, and the SN curve test results.

As part of this thesis, a Python script was developed to automate the data cleaning process. The script processes JSON files obtained from the Azure tool and converts them into structured CSV files. Approximately 100 features were initially extracted from the dataset. However, many of these features contained insufficient data. Features with more than 40% missing values were removed from the dataset. Nevertheless, key features, as discussed in Chapter 2.1.3, contained sufficient data for analysis. With the aid of domain knowledge and prior research references [10, 8], the dataset was refined to include only the most relevant features, which will be used for model training.

## Challenges

- Due to an offset in page orientation, approximately 80% of the data was extracted correctly. However, the remaining 20% was either missing or misinterpreted. This issue was manually corrected for the entire aluminum dataset and for a few data points in the steel dataset.
- Due to variations in process features, data from different sources could not be merged into a unified dataset.

## 3.3 Features and labels

With reference to previous researches[10, 8] and FKM method [11], features were selected for model training. The features are primarily categorized into process features and engineering features, which are further subdivided into chemical composition, heat treatment for process features, and mechanical properties, component design, load, and testing features as engineering features. Each subcategory contains several individual features.

Category	Feature name	Description	Derived from
Mechanical Properties	$R_m$	Tensile stress (Mpa)	Experiments
	$S_{d,FKM}$	FKM approximated fatigue limit (MPa)	Analytical approximation
Component Design	$K_t$	Stress concentration factor	Experiments
	$G_0$	Relative stress gradient at hot spot	Analytical approximation
Load and Testing	$L_{type}$	Type of mechanical loading	Experiments
	$R_{ratio}$	Ratio of min and max stress of cyclic loading	Experiments

Table 3.2: Feature description

Table 3.2 and Table 3.3 provide an overview of the features and target variables used for data-driven modeling. As discussed in Chapter 2.1.3, the factors influencing fatigue life are considered as features. Among

Category	Label name	Description	Derived from
SN curve parameters	$S_d$	Fatigue limit (MPa)	Experiments
	$N_e$	Knee point(number of cycles)	Experiments
	K	Slope of line in finite region	Experiments
	$T_s$	Scatter in infinite region	Experiments

Table 3.3: Target label description

these, only  $G_0$  and  $S_{d,FKM}$  are derived analytically, as demonstrated in Chapter 2.1.3, while all other features are directly obtained from the book. The target labels, which represent SN curve using four parameters, are estimated (refer to Chapter 2.1.2).

Due to variations in data sources and the lack of comprehensive information, process-related features such as **chemical composition and heat treatment have been excluded from the feature set**. Each database employs different material processing methods, making it challenging to standardize these features across datasets. Therefore, the development of the data-driven model focuses exclusively on engineering features.

## 3.4 Data analysis

This section presents the data analysis of key features and target variables using pair plots, heat maps, and correlation metrics. The analyzed features include tensile strength ( $R_m$ ), stress concentration factor ( $K_t$ ), and fatigue limit ( $S_d$ ) for both steel and aluminum.

$R_m$ ,  $K_t$ , stress ratio ( $R_{ratio}$ ), and stress gradient ( $G_0$ ) are numeric features, and load type ( $L_{type}$ ) is a categorical characteristic.

The green data points in Figure 3.1 represents NIMS + RB dataset, while the blue and orange points correspond to JSMS and DaBef datasets, respectively. It is evident from the figure that the majority of the data

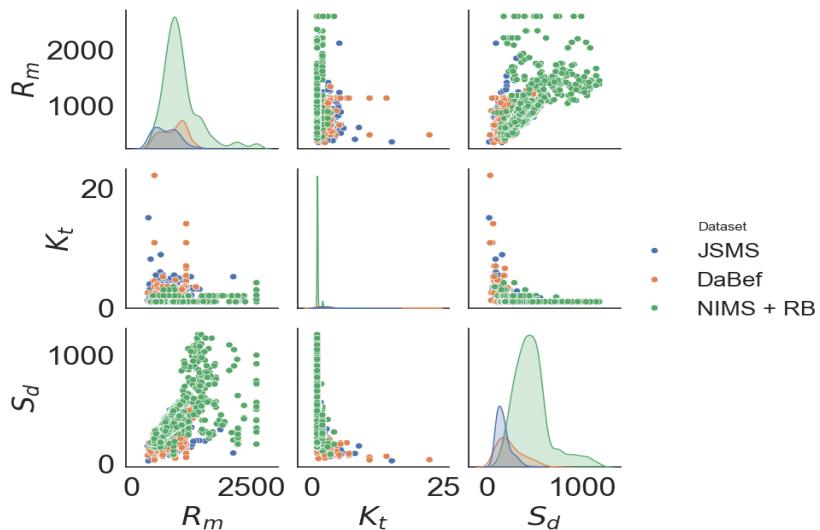


Figure 3.1: Steel - Pair plot of  $R_m$ ,  $K_t$  and  $S_d$

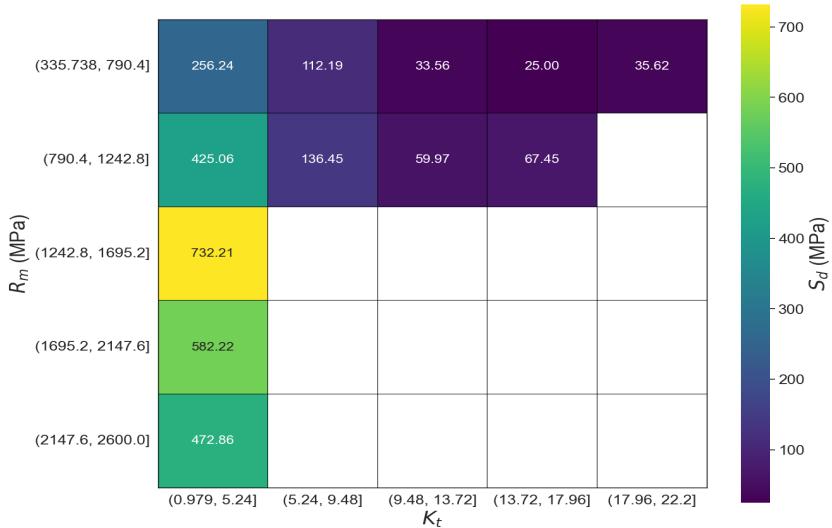


Figure 3.2: Steel - Heat map of  $R_m$ ,  $K_t$  and  $S_d$

originates from the NIMS + RB source (refer to Table 3.1). However, this dataset lacks variability, as most data points correspond to smooth specimens ( $K_t = 1$ ), which is reflected in the prominent green spike in the probability density estimate (PDE) plot of  $K_t$ . In contrast, the JSMS and DaBef datasets predominantly contain notched specimens ( $K_t > 1$ ), as indicated by the presence of orange and blue points in the  $K_t$  scatter plots.

Furthermore, it can be observed that notched specimens from the JSMS and DaBef datasets exhibit lower tensile strength ( $R_m$ ) and fatigue strength ( $S_d$ ), as indicated by the lower mean values of the blue and orange distributions in the PDE plots of  $R_m$  and  $S_d$ , compared to the green distribution.

As discussed in Chapter 2.1.3, the influence of  $R_m$  and  $K_t$  on  $S_d$  is depicted in Figure 3.2. The heat map shows that  $S_d$  increases with increasing  $R_m$  (y-axis) but decreases with increasing  $K_t$  (x-axis). Additionally, it can be observed that the steel dataset contains data points within the range of 300 MPa to 2600 MPa for  $R_m$  and 1 to 22 for  $K_t$ , indicating a diverse dataset when all three data sources are combined. However, some regions lack high  $K_t$  ( $>5$ ) data points for  $R_m$  values exceeding 1200 MPa. According to research [7], having high  $K_t$  with higher tensile strength ( $R_m$ ) is an unfeasible combination in real-world applications.

However, for steel, it is observed that beyond a certain threshold of approximately  $R_m = 1500$  MPa, a further increase in  $R_m$  results in a decrease in  $S_d$ . This phenomenon occurs due to a fundamental shift in the fracture mechanism, where specimen failure transitions from surface initiation to **internal volume fracture**. According to research findings [7], this abnormal fatigue fracture behavior is observed at  $R_m > 1400$  MPa.

From Figure 3.3, it is evident that the aluminum dataset is significantly smaller compared to the steel dataset. The aluminum dataset, derived from sources like JSMS and DaBef, exhibits a wide variability with  $R_m$

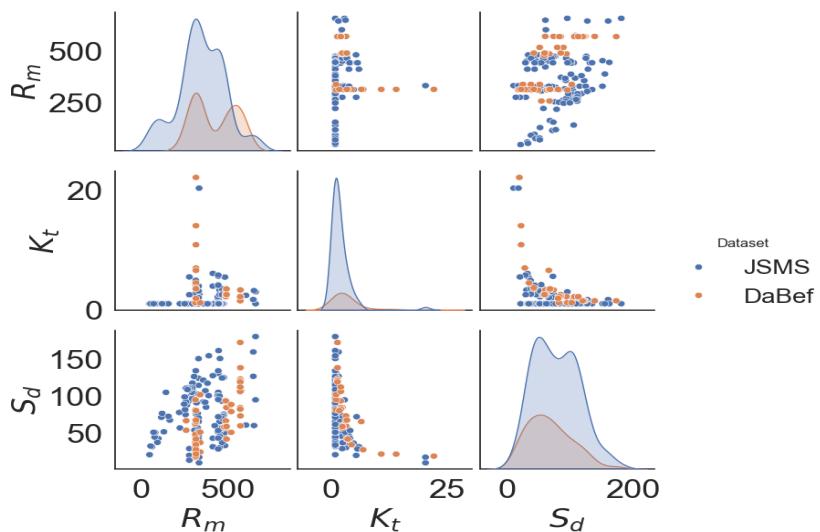


Figure 3.3: Aluminum - Comparison of  $R_m$ ,  $K_t$  and  $S_d$

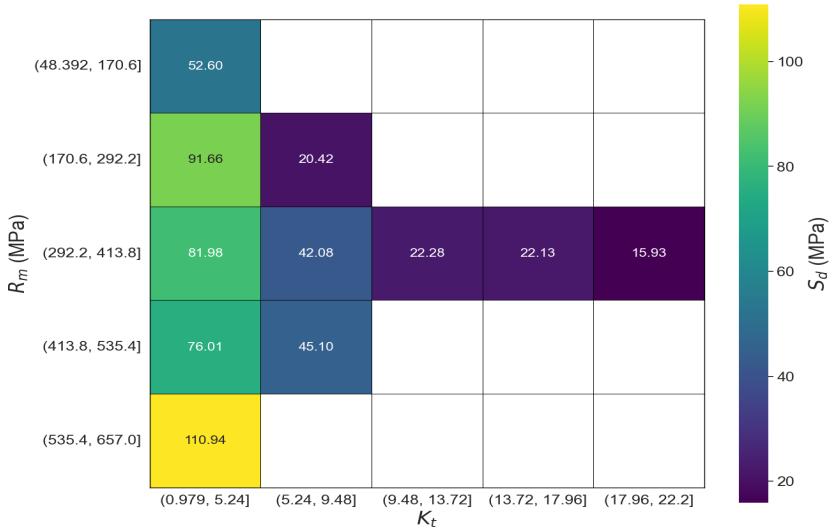


Figure 3.4: Aluminum - Heat map of  $R_m$ ,  $K_t$  and  $S_d$

ranging from 45 MPa to 650 MPa, and includes a sufficient number of notched specimens with  $K_t$  values ranging from 1 to 22. As discussed in Chapter 2, aluminum components typically exhibit lower static and fatigue strength compared to steel, which results in a maximum  $R_m$  of 657 MPa and a maximum  $S_d$  of 178 MPa.

From Figure 3.4, it can be observed that certain regions are lacking data, such as for  $R_m$  values below 413 MPa and  $K_t$  values greater than 5. These are the feasible regions, as an increase in  $K_t$  generally results in lower  $R_m$  and  $S_d$ .

However, despite these differences, the fatigue damage mechanism in aluminum follows similar patterns to steel. As shown in Figure 3.4, higher  $R_m$  values lead to higher  $S_d$  values, while higher  $K_t$  values correspond to lower  $S_d$  values.

The absence of **abnormal volumetric crack initiation** in aluminum can be attributed to its lower tensile strength and more ductile properties in comparison to steel, which is more brittle at high strength levels.

## Features and target correlation

Figure 3.5 illustrates the correlation of features and fatigue limit for aluminum using the Pearson correlation coefficient, given by:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

The Pearson correlation coefficient  $r$  ranges from  $-1$  to  $1$ :

- $r = 1$  indicates a perfect positive linear correlation.
- $r = -1$  indicates a perfect negative linear correlation.
- $r = 0$  indicates no linear relationship.

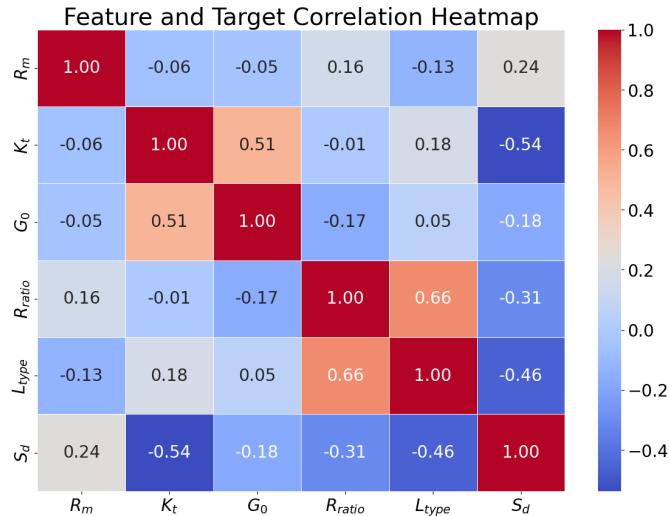


Figure 3.5: Features and target correlation heat map

From the figure, it can be observed that tensile strength ( $R_m$ ) has a positive correlation of 0.24 with the fatigue limit ( $S_d$ ), whereas the stress concentration factor ( $K_t$ ) has a negative correlation of 0.54 with  $S_d$ . This behavior aligns with expected physical equations and relationships (refer to Chapter 2.1.3).

However, there is a strong correlation between features such as  $K_t$  and the stress gradient ( $G_0$ ), as well as between the stress ratio ( $R_{ratio}$ ) and the loading type ( $L_{type}$ ). The correlation between  $R_{ratio}$  and  $L_{type}$  is justified due to the limited available data, which includes only tension-compression and bending load types. Theoretically, there is no strong correlation between  $G_0$  and  $R_{ratio}$ , as they represent different characteristic information. However, since  $G_0$  is analytically calculated and the availability of aluminum data is limited, this correlation could still hold validity.

# 4 Approaches

This chapter primarily explores data splitting techniques suitable for scenarios with limited data availability. It also examines various training methodologies and feature sets used to develop a predictive model for the fatigue limit of aluminum specimens. The combination of training methods and feature sets is systematically evaluated using a structured design of experiments approach. Finally, the chapter concludes with a comparative analysis of the results obtained from different methods and addresses key data-related questions.

## 4.1 Data split and testing methods

In this section, various data splitting techniques and their outcomes are analyzed based on the performance of Random Forest (RF) and Neural Network (NN) models. All the techniques considered use default hyperparameters (refer to Appendix B) and Feature Set 1 (refer to Table 4.3).

As discussed in Chapter 2, the standard practice initially involved using a random data split with a 70:15:15 ratio. However, the aluminum dataset contains very few data points and exhibits significant heterogeneity, making this approach challenging. The test data may not be representative of the training and validation sets, leading to inconsistencies in model evaluation. Additionally, this method is highly sensitive

to the data split seed, as different seeds yield varying accuracy levels, reducing the reliability of the results.

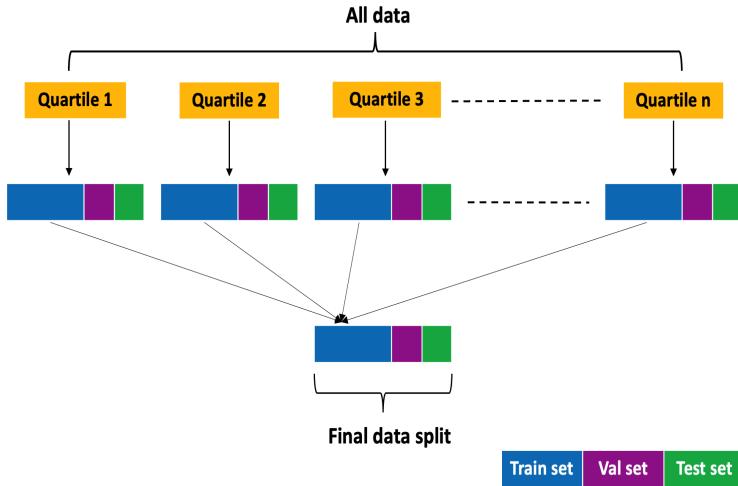


Figure 4.1: Quartile based random split

To mitigate the issue of seed sensitivity in the random split method, **quartile-based random split** strategies were implemented. As illustrated in Figure 4.1, the entire dataset was initially divided into four quartiles  $[0, 0.25, 0.5, 0.75, 1]$  based on the target variable. Within each quartile, data was randomly split into training, validation, and testing subsets in 70:15:15 ratio, respectively. The subsets from all quartiles were then merged to form a single training, validation, and testing dataset. This approach aimed to maintain a similar distribution of the target variable across all subsets. However, the method did not produce satisfactory results due to the very limited number of data points in the boundary quartiles.

Table 4.1 presents the performance of deterministic NN and RF models using different data splitting methods under varying seed values. Notably, in the random split method with seed 5, both models exhibit

Data Split Method	Seed	Models	MAE	R <sup>2</sup>
Random Split	5	NN	16.07	0.74
		RF	19.07	0.68
	32	NN	8.97	0.85
		RF	7.6	0.88
Quartile-Based Split	5	NN	14.06	0.74
		RF	12.99	0.72
	32	NN	8.99	0.87
		RF	10.05	0.78

Table 4.1: Seed Sensitivity

significantly higher MAE compared to seed 32. When evaluating performance based on MAE, NN outperforms RF for seed 5, whereas RF demonstrates superior performance for seed 32.

Furthermore, implementing the quartile-based data split method reduced the variance in predictions caused by seeding. However, seed sensitivity persists, as RF performs best with seed 5, while the NN model achieves the highest performance with seed 32. Such variations in model performance across different seed values complicate the evaluation process and model selection. Therefore, to ensure generalization and consistent results, it is essential to evaluate the entire dataset comprehensively.

## K-fold testing

To overcome the challenges of random split and quartile based split methods. K-fold testing approach has been used in this thesis. This method eliminates the sensitivity of predictions to seeding, as the entire dataset is utilized for testing during the evaluation process.

The K-fold testing is illustrated in Figure 4.2. This approach, which outperforms traditional data-splitting techniques, begins by dividing the dataset into  $k$  folds, ensuring an equal number of data points in each

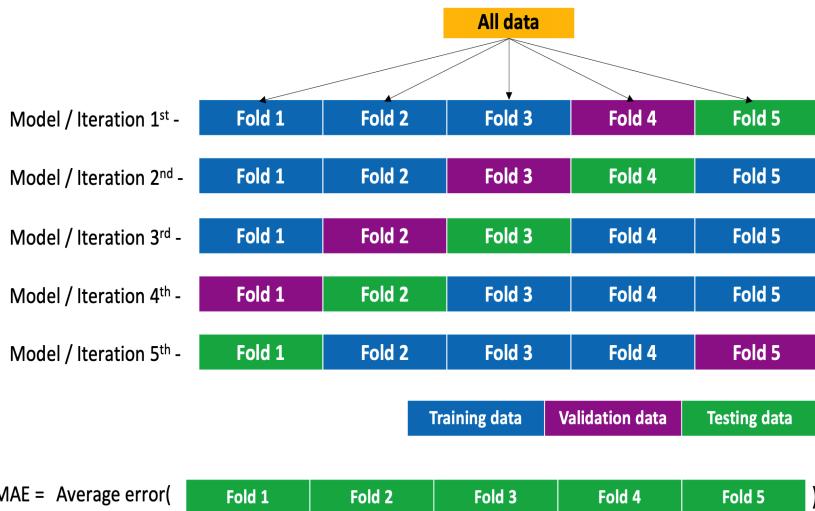


Figure 4.2: K-fold testing method

fold. In each iteration, one fold serves as validation data, another as testing data, and the remaining folds as training data.

The models are trained using the respective training and validation datasets, followed by predictions on the test dataset. The process is then repeated  $k$  times, with the models having same hyperparameters (Appendix B) being reinitialized in each iteration to prevent bias and ensure that the testing data remains unseen. This iterative procedure covers the entire dataset over  $k$  iterations, resulting in predictions for all data points. Finally, these predictions are used to evaluate the models using metrics such as MAE,  $R^2$  score, and NLL.

The advantage of the  $k$ -fold testing method over random splitting is illustrated in Figure 4.3. It depicts the MAE distribution for random forest and neural network models, observed using 15 different seeds for splitting the data with both random and  $k$ -fold techniques.

The blue density plots represent the MAE observed for the random split,

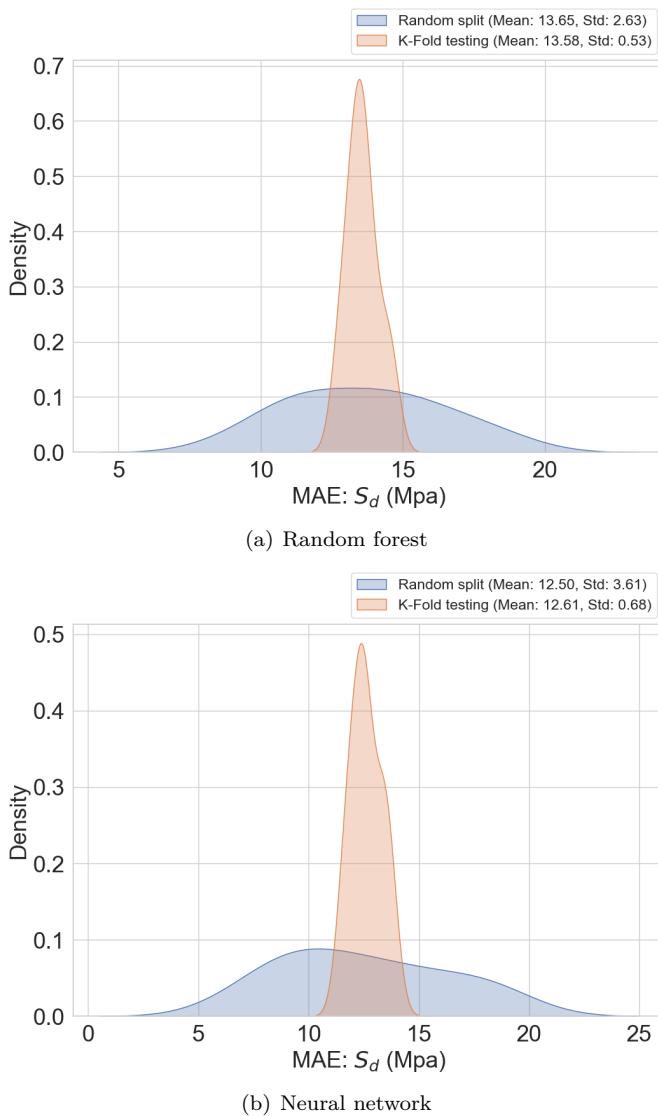


Figure 4.3: Data split method comparison

while the orange density plots correspond to the  $k$ -fold method. The standard deviation of the blue plots is significantly higher compared to the orange plots, making evaluation challenging. This highlights the sensitivity of the random split technique to seeding, affecting predictions and model evaluation. In contrast, the  $k$ -fold testing method demonstrates lower variance in its predictions, making it seeding-insensitive. Consequently, the  $k$ -fold evaluation technique is used for further model evaluation in this thesis.

## Bayesian hyperparameter optimization

In the previous section, we observed the issues associated with using random split and quartile-based data split. Therefore, it was concluded to use the entire dataset as testing data through the  $k$ -fold testing method. However, this approach also presents challenges for hyperparameter optimization, as test data needs to be kept aside for conventional  $k$ -fold evaluation.

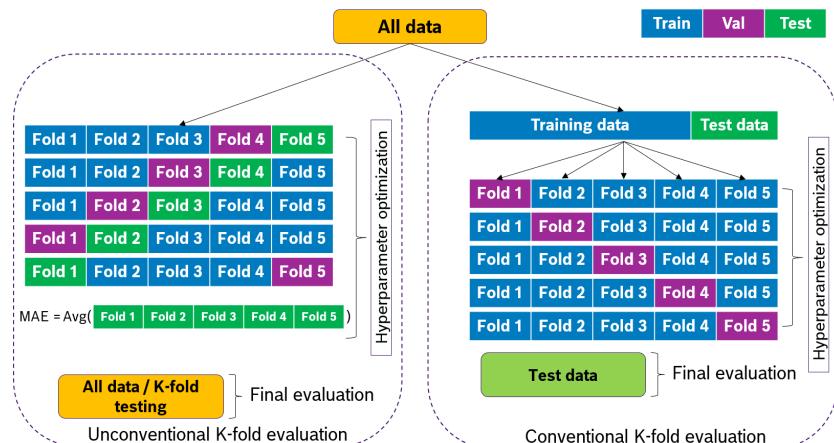


Figure 4.4: Conventional and Unconventional K-fold evaluation

To address this problem, an unconventional k-fold evaluation technique was explored, as illustrated in Figure 4.4. Unlike the conventional approach, where the test dataset is excluded during hyperparameter tuning, this method utilizes the entire dataset for optimization. The same k-fold testing strategy is then applied for final testing with different split of folds for avoiding bias. This approach aims to make the selected hyperparameters more robust, leading to better generalization.

However, using the entire dataset for hyperparameter optimization introduces biased parameters, as the model has access to all data during tuning. This method was primarily implemented to assess the potential performance of the models under optimal conditions.

Category	Models	MAE	R <sup>2</sup>	NLL
Non tuned models	RF	12.92	0.77	
	SVR	14.70	0.72	
	KNN	20.75	0.50	
	NN	12.19	0.81	
	GPR	11.37	0.82	5.63
	RFCI	12.86	0.77	7.65
	PNN	11.86	0.81	3.21
	Deep ensembles	11.64	0.83	3.23
Hyperparameter tuned models	RF	12.12	0.79	
	SVR	13.99	0.75	
	KNN	19.44	0.54	
	NN	11.14	0.84	
	GPR	11.06	0.82	3.73
	RFCI	12.13	0.78	6.43
	PNN	10.65	0.83	3.13
	Deep ensembles	11.15	0.83	3.14

Table 4.2: Aluminum results – Comparison of Hyperparameter-Tuned and Non-Tuned Models

Despite this, it was observed that hyperparameter tuned models did not show significant improvement compared to non optimized models (Table 4.2). This phenomenon occurs due to the low complexity of the models. This is further supported by the fact that a linear regression model also performs well to some extent (refer to Chapter 4.3). Additionally, as

seen in Table 4.2, less complex models like K Nearest Neighbor (KNN) also demonstrate the ability to model the data to some extent. Additionally, Mantovani et al. [17] proposed a meta-learning recommender system to predict when hyperparameter tuning would significantly improve model performance, highlighting that tuning does not always yield substantial benefits.

Therefore, the development of models to determine the most suitable methodology and approach for predicting fatigue parameters was conducted without hyperparameter optimization. Based on prior research [8, 4] and past experience, standard hyperparameters were used for this thesis (refer to Appendix B).

## 4.2 Methodologies

This section outlines various methodologies, including three distinct training strategies with different feature set combinations. These strategies focus on different approaches to develop Machine learning (ML) models for predicting fatigue parameters of steel and aluminum. The optimal combination of training strategy and feature set is determined through a series of experiments. Furthermore, this section also discusses ML model selection and strategies to avoid over fitting of models.

### Training strategies

Figure 4.5 illustrates the different methods of training strategies. The first method involves using **separate models** for steel and aluminum. These models are trained separately with their respective steel and aluminum datasets.

The second method explores the possibility of a **combined model** for predicting fatigue parameters of both steel and aluminum together. In

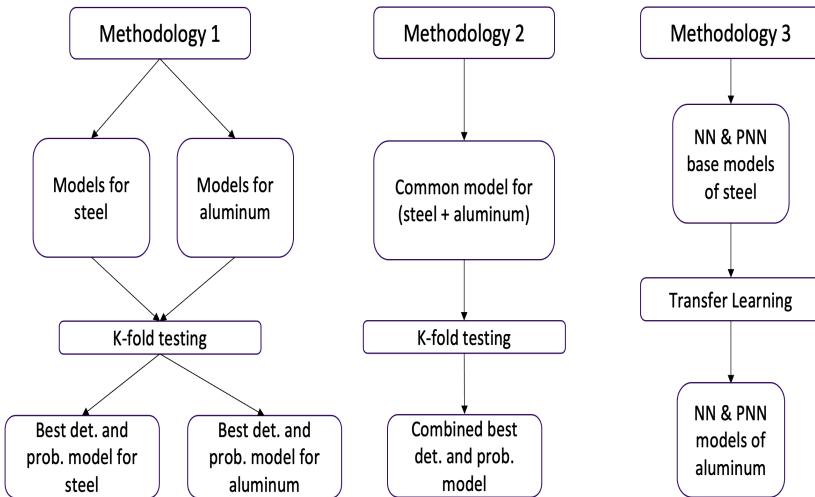


Figure 4.5: Different training strategies

this approach, an additional feature that classifies the material as steel and aluminum is added to the feature set. This method identifies the best deterministic and probabilistic models capable of predicting fatigue parameters for both materials.

Finally, the third method utilizes the **transfer learning** technique. In this method, a trained steel model is used to develop the aluminum model. This approach is applicable only for Neural Network (NN) based models. During training, all layers of the trained NN, except the final layer, are frozen. The model is then fine-tuned with aluminum data by training the weights of the last layer, allowing it to make predictions for aluminum. This technique leverages the knowledge learned from the steel model, stored as weights in the frozen layers, to improve predictions for aluminum.

## Feature sets

Category	Feature name	FKM	Set 1	Set 2	Set 3
Experimental features	$R_m$	✓	✓	✓	✓
	$K_t$	✓	✓	✓	✓
	$L_{type}$	✓	✓	✓	✓
	$R_{ratio}$	✓	✓	✓	✓
Analytical derived features	$G_0$	✓	✓	✓	
	$S_{d,FKM}$			✓	

Table 4.3: Feature sets

With domain expertise and prior knowledge from past research studies [9], features are selected based on their contribution to estimate fatigue parameters. Table 4.3 presents the various combinations of features used for model training, which are primarily categorized into experimental and analytical features.

The FKM set represents the features utilized in the FKM analytical approach (refer to Chapter 2.1.4). Since the FKM method serves as a benchmark, the first feature set for data driven model is designed to closely align with the FKM features to enable direct comparison.

The second feature set extends this by incorporating an additional fatigue limit ( $S_{d,FKM}$ ) feature derived from the FKM analytical approach (Figure 2.10), which is expected to enhance model performance. This method acts as an error correction technique, as  $S_{d,FKM}$  is expected to be close to the actual fatigue limit ( $S_d$ ).

Meanwhile, the third feature set includes only experimental features to explore the feasibility of **black box modeling**, as only experimental features will be used. However, from an engineering standpoint, stress gradient ( $G_0$ ) provides crucial insights into the stress distribution at the notch and significantly influences fatigue limit estimation.

## Design of experiments

DoE No.	Training method			Feature Sets		
	First Method	Second Method	Third Method	Set 1	Set 2	Set 3
DoE 1	✓			✓		
DoE 2	✓				✓	
DoE 3	✓					✓
DoE 4		✓				✓
DoE 5			✓	✓		✓

Table 4.4: Design of experiments

The combinations presented in Table 4.4 illustrate the various Design of Experiments (DoE) configurations employed to identify the most suitable approach for data-driven modeling of fatigue parameter predictions. The objective of these experiments is to systematically evaluate different feature sets and training methods to determine the optimal strategy for accurate fatigue parameters estimation.

Initially, the first three DoE configurations were designed to assess the influence of different feature combinations, aiming to identify the best possible feature set for fatigue parameter prediction. Subsequently, two additional DoE configurations were introduced to compare different training methods while maintaining a consistent feature set.

The first two DoE configurations adopt a **hybrid modeling** approach, where analytical features derived from the FKM analytical method are incorporated as input parameters alongside experimental features. This allows for a direct comparison between traditional analytical modeling and data-driven approaches. In contrast, the last three DoE configurations are specifically designed to evaluate the performance of purely **black-box models**, relying solely on experimental features without any analytical feature inputs.

Each DoE configuration is applied to datasets from both steel and aluminum. However, due to the complexity involved in computing analytical features, some datasets do not currently include these features. In particular, the DaBef and JSMS datasets for steel lack the required analytical features at this stage. On the other hand, the analytical features for the aluminum dataset are fully available across the entire dataset.

To ensure a fair comparison, the DoE evaluations are conducted as follows:

- DoE 1 and 2: Since these experiments require analytical features, they are applied to the complete aluminum dataset and a partial steel dataset comprising the NIMS and RB datasets only.
- DoE 3: Since this DoE relies only on experimental features, it is evaluated using two different steel dataset configurations: (1) a partial dataset (NIMS + RB) and (2) the full dataset (NIMS + RB + JSMS + DaBef). The purpose of this dual evaluation is to analyze the influence of additional data, particularly datasets with a higher proportion of notched specimens.
- DoE 4 and 5: These experiments also require only experimental features. Therefore, they are conducted using the entire steel and aluminum dataset, ensuring a comprehensive evaluation of black box modeling performance across all available data.

## Separate models for each fatigue parameters

Due to the highly nonlinear nature of fatigue parameters such as knee point ( $N_e$ ), slope in finite region ( $K$ ) and scatter in infinite region ( $T_s$ ), developing a single machine learning model to predict all of them accurately is challenging. These parameters introduce more errors compared to the prediction of fatigue limit ( $S_d$ ).

Previous research [8] has shown that a single model for all four fatigue parameters results in degraded performance. Therefore, this thesis focuses on developing separate machine learning models for each fatigue parameter.

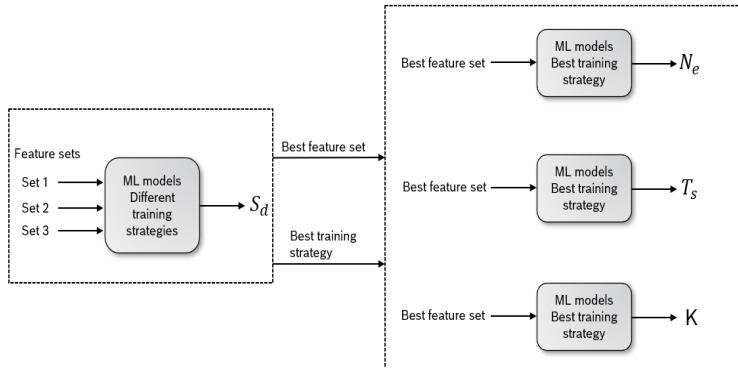


Figure 4.6: Separate models for fatigue parameters

As illustrated in Figure 4.6, different combinations of feature sets and training strategies are evaluated using design of experiments. The best feature set and best training strategy will then be applied to the other fatigue parameter models.

## Machine learning models

Due to the limited availability of aluminum data, selecting suitable Machine Learning (ML) models presents a challenge. Therefore, several commonly used ML models are considered in this thesis. Deterministic ML models such as Linear Regression (LR), K-Nearest Neighbors (KNN), and Support Vector Regressor (SVR) are employed due to their lower complexity and serve as a baseline for comparing the complexity of physical models with more advanced ML models like Random Forest

(RF) and Neural Networks (NN). Additionally, the LR model helps quantify the linearity and basic complexity of physics-based models. Given the highly complex nature of the fatigue mechanism, as discussed in Chapter 2.1.1, nonlinear models like RF and NN are considered, based on prior research [4, 8], where RF was identified as the best-performing deterministic model for steel.

Furthermore, probabilistic models such as Random Forest with Confidence Interval (RFCI), Probabilistic Neural Network (PNN), and Probabilistic Deep Ensembles have been used in previous research [8] for steel due to the necessity of uncertainty quantification. Additionally, research [10] has employed Gaussian Process Regressor (GPR), which performs well with limited data. Since the aluminum dataset consists of only 130 data points (Table 3.1), GPR is considered an ideal choice among probabilistic ML models.

### **Why is uncertainty quantification required?**

Due to the small and highly heterogeneous aluminum dataset, as discussed in Chapter 3.4, assessing the confidence of ML model predictions is crucial. Probabilistic ML models provide variance along with predictions, allowing engineers to determine the reliability of the predictions. ML models typically yield higher variance for data points with larger absolute errors.

To determine the best models, deterministic and probabilistic ML models are trained and evaluated with same dataset. All models are trained using default hyperparameters (see Appendix B), and predictions are made on the entire dataset using a k-fold testing technique (Figure 4.2). The best deterministic and probabilistic models are then identified by comparing the error metrics across all models.

## Avoiding Over-fitting

Over-fitting occurs when a model learns the noise in the training data rather than the underlying pattern, leading to poor generalization on unseen data. Different models have varying susceptibility to over-fitting, and multiple techniques can mitigate this issue. Simpler models like Linear Regression (LR) and K-Nearest Neighbors (KNN) have a lower chance of over-fitting. Support Vector Regression (SVR) over-fitting can be avoided through appropriate kernel selection and tuning of the regularization parameter  $C$  and the  $\epsilon$ -insensitive loss function. Similarly, over-fitting in Gaussian Process Regression (GPR) can be mitigated using proper kernel selection and noise regularization. Neural Networks (NN), on the other hand, are highly prone to over-fitting due to their high degree of freedom.

- **Random Forest (RF):** Controlling tree depth and limiting the number of leaves help in reducing over-fitting.
- **Neural Networks (NN):** Techniques such as dropout, batch normalization, and early stopping help prevent over-fitting in deep learning models.

Regularization and other techniques to prevent over-fitting in various ML models are detailed in Appendix B.

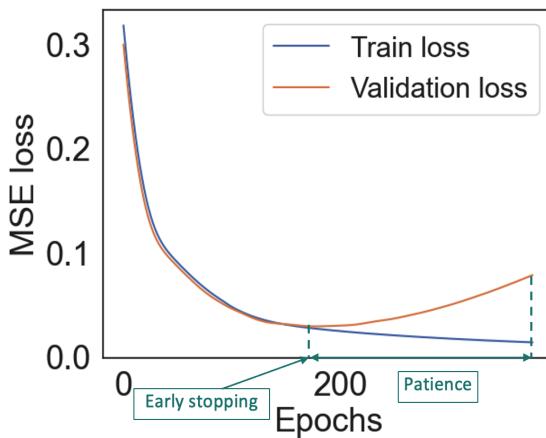


Figure 4.7: Early stopping method

Figure 4.7 shows training and validation loss curve of deterministic NN. In this approach, the training of the neural network is halted as soon as there is no improvement in the validation loss or if the loss starts to increase for a certain number of epochs, referred to as patience. Once training is stopped, the model weights corresponding to the lowest validation loss are reassigned to the network to ensure optimal performance.

## 4.3 Results

The results discussed in this section represent the best outcomes from both deterministic and probabilistic Machine Learning (ML) models for fatigue limit ( $S_d$ ) predictions of aluminum across all DoE. Additionally, plots illustrating other fatigue parameters, such as knee point ( $N_e$ ), slope in the finite region ( $K$ ), and scatter in the infinite region ( $T_s$ ) are shown in Appendix A.

Mainly, this section presents a scatter plot for predictions, a box plot for the error of predictions, and tabular model performance metrics. In the scatter plot of predictions, the x-axis represents the target values, while the y-axis denotes the corresponding predictions. The dark dashed line indicates the ideal prediction line, whereas the parallel dotted lines mark the 10% tolerance limits (i.e., 10% of the range). Additionally, the plot for probabilistic models includes the 95% confidence interval lines.

The box plot provides a visual representation of the distribution of errors. The y-axis represents the actual error in predictions, while the x-axis represents all models. Similar to the scatter plot, this visualization includes an ideal horizontal line at zero, along with parallel lines marking the 10% tolerance limits (i.e., 10% of the range). It highlights key statistical measures such as the median, interquartile range (IQR), whiskers, outliers, and the mean. The median, represented by a central line within the box, indicates the middle value of the all the errors, while the IQR captures the spread of the middle 50% of the error. The whiskers extend to the smallest and largest values within 1.5 times the IQR, and any error points beyond this range are considered outliers, marked as individual points. Additionally, the inclusion of the mean, represented by a marker, provides another measure of central tendency.

Furthermore, a table of model performance metrics showcases parameters such as Mean Absolute Error (MAE),  $R^2$  score, and Negative Log-Likelihood (NLL) for model comparison.

### 4.3.1 DoE 1

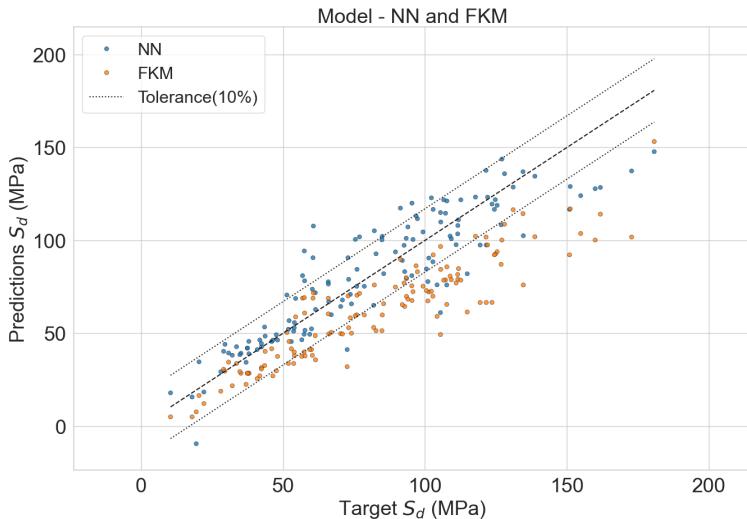


Figure 4.8: DoE 1 - Neural Network (Aluminum)

DoE 1 represents the separate model training method with feature set 1, which replicates the FKM features (refer to Chapter 4.2). Figure 4.8 shows the comparison of predictions made by the Neural Network (NN), which performed best among the deterministic ML models, and the predictions obtained using the FKM analytical method (Figure 2.10).

It is evident from the figure that the NN outperformed the FKM method, as the majority of predictions align closely with the ideal line and fall within the tolerance limits. However, the FKM method is inherently conservative, leading to under-predicted results. In contrast, while the NN aims for a mean error of zero, some predictions are overly optimistic, as it does not incorporate conservatism in its predictions.

Among the probabilistic models, the Probabilistic Neural Network (PNN) performed the best, achieving the lowest NLL and MAE. Figure 4.9 shows the scatter plot of predictions made by the PNN model.

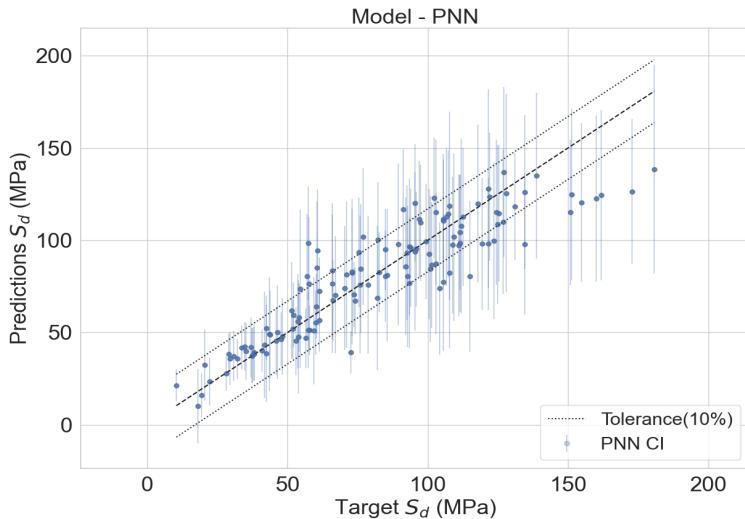


Figure 4.9: DoE 1 - Probabilistic Neural Network (Aluminum)

Similar to the NN model, it can be observed that the majority of predictions align with the ideal line and fall within the tolerance limits. Additionally, predictions that deviate significantly from the ideal line, exhibiting higher errors, also have higher variance. This indicates that the worst predictions made by the PNN model are associated with greater uncertainty, highlighting their lower reliability from an engineering perspective.

Both NN and PNN models exhibit high prediction errors for target values above 150 MPa due to the limited availability of data in this range.

As observed in Figure 4.10, the error distribution plot shows that the Random Forest (RF) and NN models have a small IQR, with most predictions falling within the tolerance limits. However, both models contain a few outliers with errors extending up to  $\pm 50$  MPa. Similarly, Gaussian Process Regressor (GPR) and PNN demonstrate strong performance among probabilistic models.

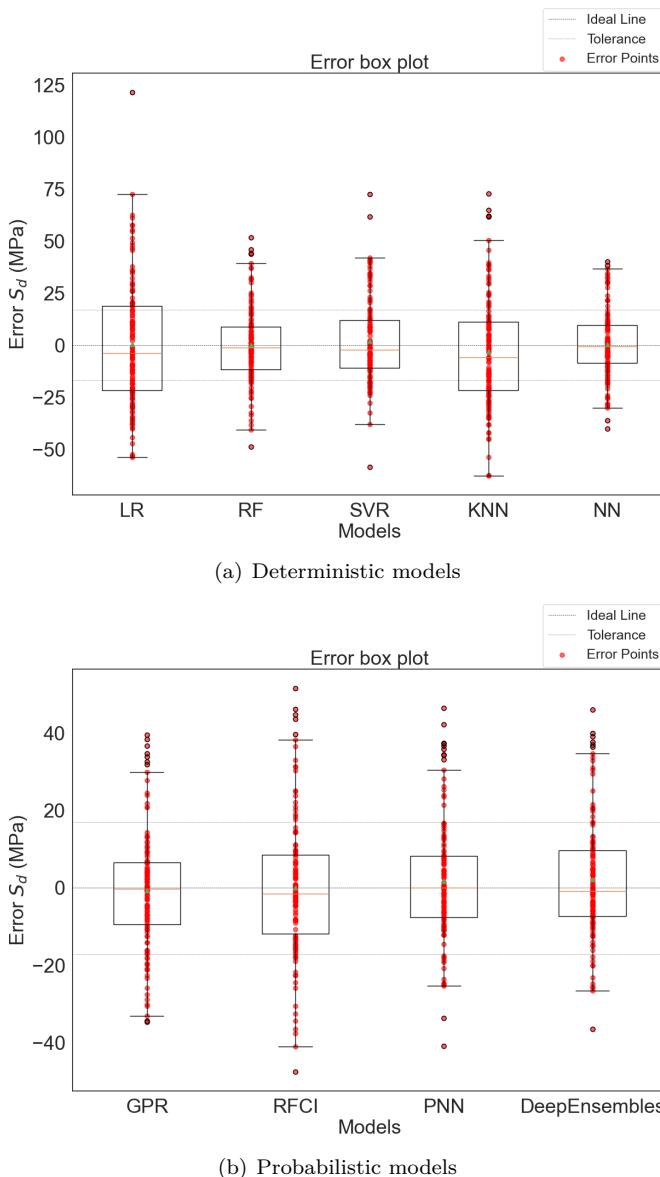


Figure 4.10: DoE 1 - Error distribution plot (Aluminum)

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
FKM analytical method	FKM	22.53	0.42	
Deterministic models	LR	23.79	0.34	
	RF	13.22	0.76	
	SVR	15.42	0.69	
	KNN	21.53	0.46	
	NN	11.76	0.81	
Probabilistic models	GPR	11.69	0.81	5.57
	RFCI	13.19	0.76	7.16
	PNN	11.49	0.82	3.17
	Deep Ensembles	11.40	0.82	3.19

Table 4.5: DoE 1 - Performance metrics (Aluminum)

Table 4.5 presents the performance metrics of all models. The FKM analytical method resulted in an MAE of 22.53 MPa and an  $R^2$  score of 0.42. In contrast, machine learning models significantly outperformed the FKM approach, with the NN achieving the lowest MAE of 11.76 MPa and the highest  $R^2$  score of 0.81. Alongside NN, Random Forest (RF) also performed equivalently well among deterministic models, with an MAE of 13.22 MPa.

LR performed the worst due to its inability to capture nonlinearity. Furthermore, the performance of LR and FKM is nearly comparable, with FKM performing slightly better. This similarity in performance can be attributed to the FKM method utilizing simplified analytical equations for fatigue limit prediction (Figure 2.10).

All probabilistic models performed well in terms of MAE; however, the Gaussian Process Regressor (GPR) and Random Forest with Confidence Interval (RFCI) yielded higher NLL values, indicating that their predictions and variance estimates were not as reliable. In contrast, PNN and Deep Ensembles achieved NLL values close to 3.2, making them better suited for probabilistic predictions.

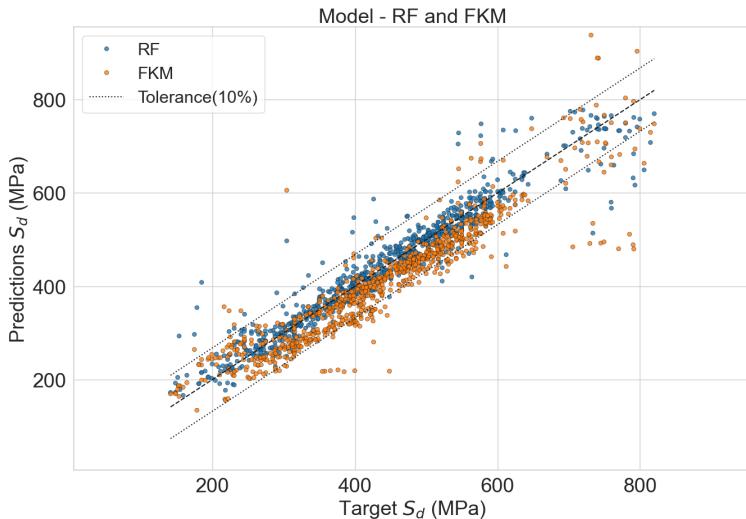


Figure 4.11: DoE 1 - Random Forest (Steel)

Category	Models	MAE	R <sup>2</sup>	NLL
FKM analytical method	FKM	44.16	0.80	
Deterministic models	LR	41.72	0.75	
	RF	23.91	0.92	
	SVR	24.31	0.91	
	KNN	24.34	0.91	
	NN	26.40	0.92	
Probabilistic models	GPR	23.48	0.93	8.46
	RFCI	23.87	0.92	7.08
	PNN	24.90	0.92	3.89
	Deep Ensembles	24.32	0.92	3.91

Table 4.6: DoE 1 - Performance metrics (Steel)

Figure 4.11 illustrates the comparison between RF model predictions and FKM method predictions for steel. Similar to aluminum, the conservative nature of the FKM method is observed, while the RF model outperforms it. However, for steel, the FKM method also produces relatively good results, as the majority of predictions (orange scatter) fall within the tolerance limits. This can be attributed to the steel data used

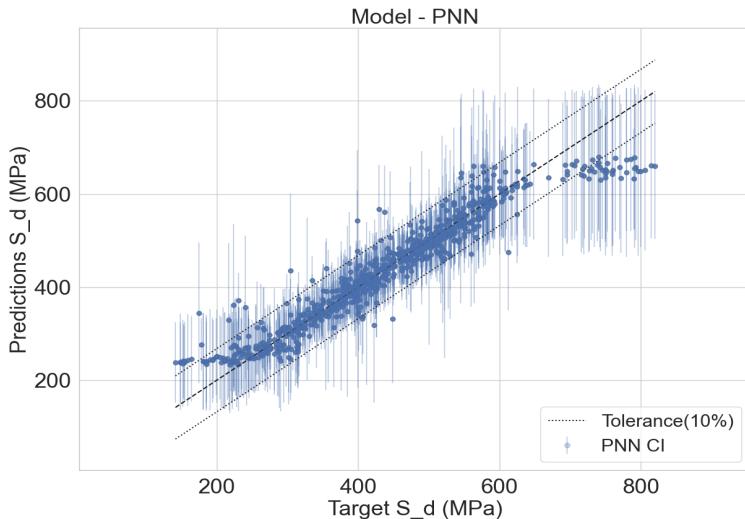


Figure 4.12: DoE 1 - Probabilistic Neural Network (Steel)

in this DoE being less complex, as it does not include many notched specimens (refer to 4.2). For the same reason, the performance of LR is also considerably good, and the performance of all models is relatively similar (Table 4.6).

In contrast to the aluminum models, RF demonstrates the best performance among deterministic models for steel, while PNN performs best among probabilistic models. From Figure 4.12, it can be observed that predictions above 700 MPa and below 200 MPa exhibit higher errors. This suggests that the PNN model is not learning effectively in these regions, likely due to the limited availability of data.

DoE 1 concludes that data-driven modeling for fatigue limit ( $S_d$ ) prediction achieves 50% lower MAE compared to the FKM method for both steel and aluminum, making it a better choice as it provides more accurate predictions.

### 4.3.2 DoE 2

In this approach, the fatigue limit calculated using the FKM method ( $S_{d,FKM}$ ) is incorporated as an additional feature to enhance model accuracy. Since  $S_{d,FKM}$  theoretically exhibits a linear relationship with the actual fatigue limit ( $S_d$ ), its inclusion improves predictive performance.

However, while the Linear Regression (LR) and K-Nearest Neighbor (KNN) models showed significant improvement, more complex models exhibited only a slight enhancement compared to DoE 1. This could be attributed to the ability of complex models to inherently capture non-linear relationships without relying on  $S_{d,FKM}$ .

Category	Models	MAE	R <sup>2</sup>	NLL
Deterministic models	LR	11.53	0.82	
	RF	11.60	0.81	
	SVR	11.20	0.83	
	KNN	14.46	0.74	
	NN	11.30	0.83	
Probabilistic models	GPR	11.15	0.83	5.86
	RFCI	11.69	0.80	8.60
	PNN	10.79	0.84	3.11
	Deep Ensembles	10.94	0.84	3.14

Table 4.7: DoE 2 - Performance metrics (Aluminum)

Category	Models	MAE	R <sup>2</sup>	NLL
Deterministic models	LR	31.49	0.86	
	RF	22.75	0.92	
	SVR	24.32	0.90	
	KNN	24.13	0.91	
	NN	25.88	0.91	
Probabilistic models	GPR	23.66	0.92	8.02
	RFCI	22.78	0.92	6.97
	PNN	23.48	0.92	3.83
	Deep Ensembles	23.73	0.92	3.90

Table 4.8: DoE 2 - Performance metrics (Steel)

### 4.3.3 DoE 3

This DoE is conducted to assess model performance using a black-box modeling approach, relying solely on experimental features. The results in Table 4.9 indicate that the exclusion of the stress gradient ( $G_0$ ) had minimal impact on the performance of aluminum models (Table 4.5). A similar trend is observed for steel, as the results of DoE 3 (Table 4.10) show no significant improvement compared to DoE 1 (Table 4.6).

The effect of  $G_0$  from a data-driven perspective may not be evident due to two reasons. First, we have used an analytical approximation of  $G_0$ , whereas ideally, an FEA-simulated  $G_0$  should be used. Second, the stress concentration factor  $K_t$  and  $G_0$  exhibit a correlation of 0.51 (Figure 3.5).

Since  $G_0$  do not significantly influence model performance, an advantage arises in utilizing the entire steel dataset, including extended notched specimen data. Furthermore, in the same DoE steel models are evaluated using an extended dataset incorporating JSMS and DaBef data. As shown in Table 4.11, the performance of all models deteriorated compared to Table 4.10. The Linear Regression (LR) in particular experienced a significant decline in performance due to the inclusion of notched specimen data, which introduces high nonlinearity in fatigue limit estimation as shown in Figure 4.14.

Neural Network (NN) and Probabilistic Neural Network (PNN) remain the best deterministic and probabilistic models, respectively, for estimating the fatigue limit of aluminum (Figure 4.13). Furthermore, for steel, Random Forest (RF) and PNN are the best deterministic and probabilistic ML models, similar to DoE 1 and DoE 2.

This DoE concludes that  $G_0$  **does not have significant importance** as a feature and can be eliminated when leveraging black-box modeling for fatigue limit prediction.

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	23.78	0.37	
	RF	12.92	0.77	
	SVR	14.70	0.72	
	KNN	20.75	0.50	
	NN	12.19	0.81	
Probabilistic models	GPR	11.37	0.82	5.63
	RFCI	12.86	0.77	7.65
	PNN	11.65	0.83	3.19
	Deep Ensembles	11.15	0.83	3.21

Table 4.9: DoE 3 - Performance metrics (Aluminum)

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	41.35	0.76	
	RF	24.68	0.91	
	SVR	24.00	0.91	
	KNN	24.60	0.91	
	NN	26.69	0.92	
Probabilistic models	GPR	23.48	0.93	8.48
	RFCI	24.77	0.91	7.38
	PNN	24.03	0.92	3.87
	Deep Ensembles	23.62	0.93	3.91

Table 4.10: DoE 3 - Performance metrics (Steel)

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	76.65	0.61	
	RF	31.28	0.92	
	SVR	35.69	0.89	
	KNN	37.18	0.86	
	NN	32.15	0.92	
Probabilistic models	GPR	31.54	0.92	15.90
	RFCI	31.27	0.92	7.14
	PNN	33.05	0.91	4.24
	Deep Ensembles	32.33	0.91	4.28

Table 4.11: DoE 3 - Performance metrics with extended data (Steel)

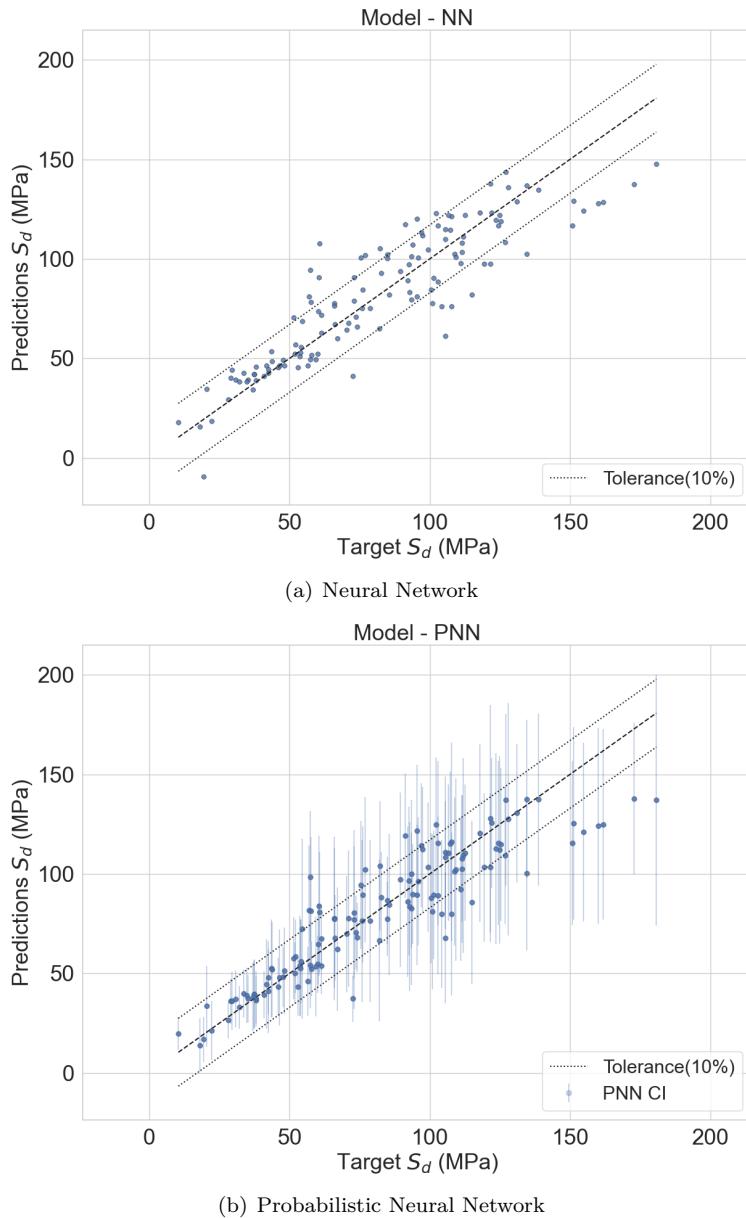


Figure 4.13: DoE 3 - Prediction plots (Aluminum)

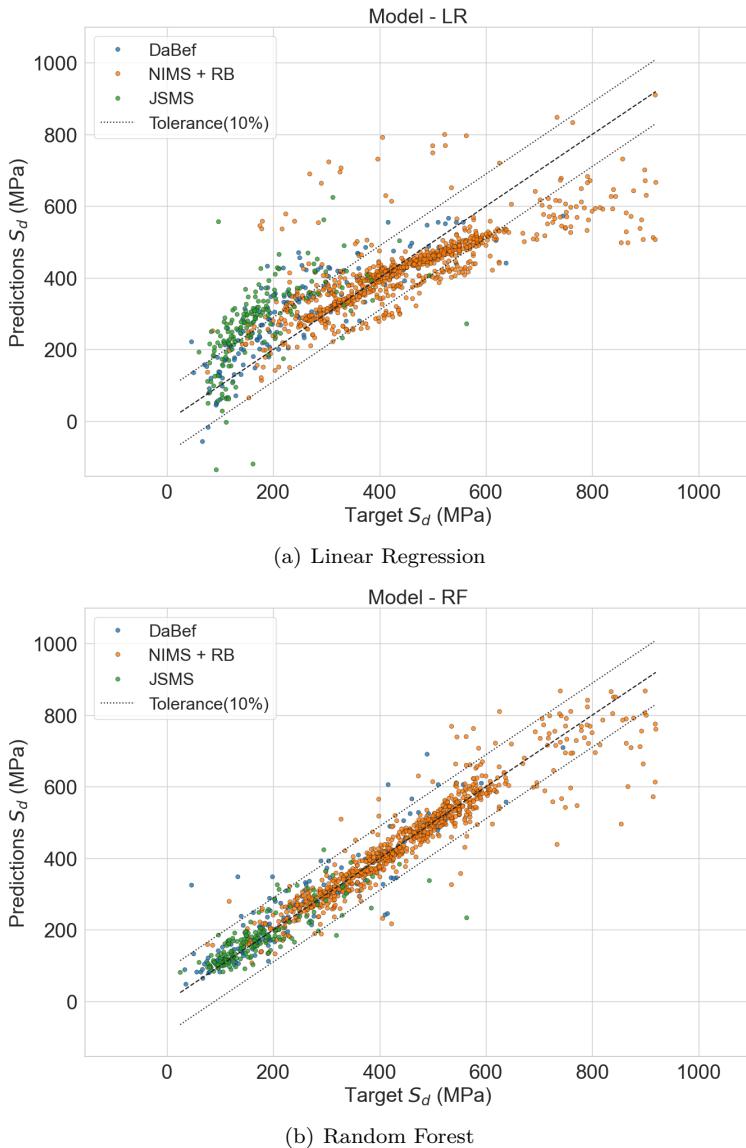


Figure 4.14: DoE 3 - Prediction plots with extended data (Steel)

### 4.3.4 DoE 4

This experiment employs a combined model for steel and aluminum, incorporating an additional feature to distinguish between the two materials. The model utilizes only experimental features, as analytical features have minimal influence on data-driven models (Concluded in DoE 3).

The overall performance of the combined model is satisfactory, with Random Forest (RF) and Neural Network (NN) has good performance among deterministic models. Among probabilistic models, Random Forest with Confidence Interval (RFCI) and Probabilistic Neural Network (PNN) demonstrate good performance. However, a decline in performance is observed specifically for aluminum predictions. As shown in Table 4.13, NN models exhibit an increased MAE and a significantly reduced  $R^2$  score compared to DoE 3 (Table 4.9). In contrast, the performance of steel remains largely unchanged when compared to DoE 3.

This discrepancy may be attributed to an **imbalanced dataset composition**, as the dataset contains significantly more steel data than aluminum data.

RF performs well in this scenario due to its ability to handle noise effectively, as it is an ensemble learning method. In contrast, models such as NN, PNN, and GPR are more sensitive to data distribution. The limited aluminum data may introduce noise for these models, leading to degraded performance.

Furthermore, from Figure 4.16, it can be observed that predictions below 50 MPa exhibit higher MAE. Additionally, the overall uncertainty of predictions has increased compared to DoE 3 (Figure 4.13).

From this DoE it is evident that this training method of combined modeling is **not good method** as it significantly deteriorated the aluminum prediction performance.

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	85.38	0.62	
	RF	32.39	0.93	
	SVR	39.64	0.88	
	KNN	37.85	0.89	
	NN	32.94	0.93	
Probabilistic models	GPR	33.16	0.92	16.98
	RFCI	32.31	0.93	8.15
	PNN	33.82	0.91	4.21
	Deep Ensembles	33.37	0.92	4.30

Table 4.12: DoE 4 - Performance metrics (Steel + Aluminum)

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	80.76	-14.35	
	RF	14.74	0.69	
	SVR	15.77	0.50	
	KNN	16.89	0.66	
	NN	18.34	0.61	
Probabilistic models	GPR	19.14	0.46	4.01
	RFCI	14.57	0.70	3.99
	PNN	13.75	0.77	3.43
	Deep Ensembles	16.22	0.70	3.67

Table 4.13: DoE 4 - Performance metrics (Aluminum)

<b>Category</b>	<b>Models</b>	<b>MAE</b>	<b>R<sup>2</sup></b>	<b>NLL</b>
Deterministic models	LR	85.88	0.58	
	RF	34.28	0.91	
	SVR	42.21	0.85	
	KNN	40.11	0.86	
	NN	34.30	0.91	
Probabilistic models	GPR	34.67	0.90	18.38
	RFCI	34.22	0.91	8.60
	PNN	35.98	0.89	4.29
	Deep Ensembles	35.21	0.90	4.37

Table 4.14: DoE 4 - Performance metrics (Steel)

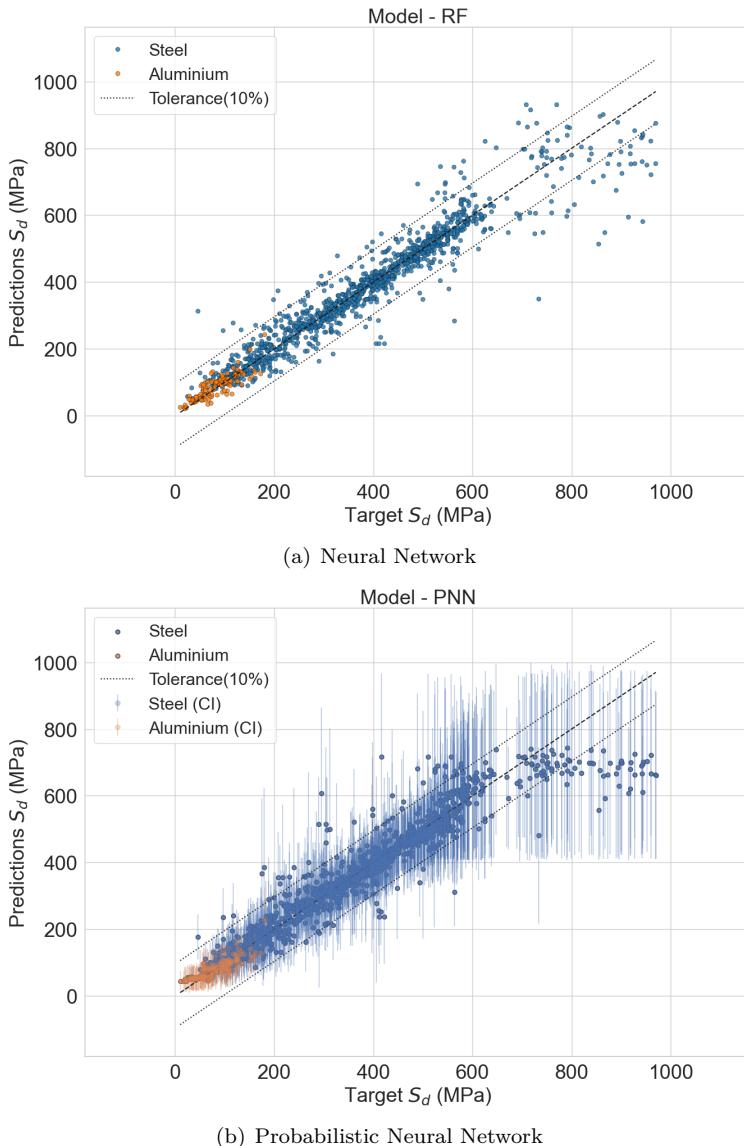


Figure 4.15: DoE 4 - Prediction plots (Steel + Aluminum)

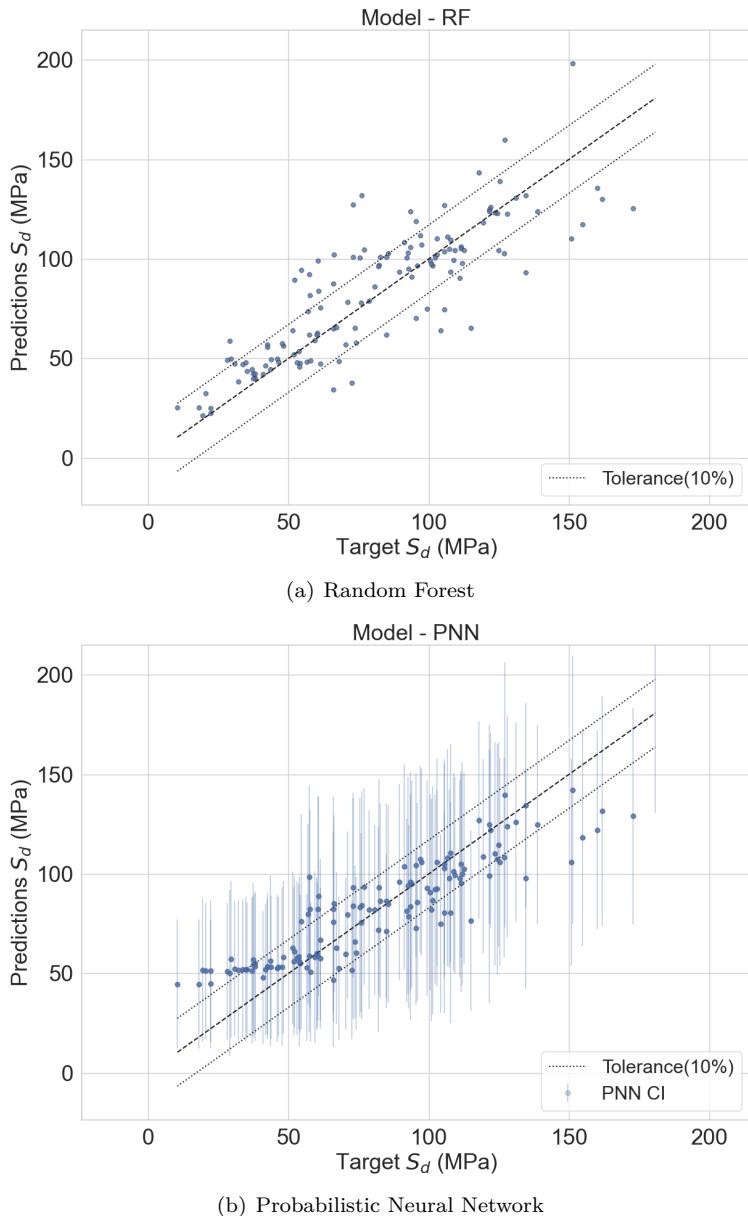


Figure 4.16: DoE 4 - Prediction plots (Aluminum)

### 4.3.5 DoE 5

The concept of transfer learning is applied to both deterministic Neural Network (NN) and Probabilistic Neural Network (PNN). A base model for steel, with the architecture shown in Appendix Table B.4, is trained using the feature set and training method from DoE 3. This trained steel model consists of three hidden layers with four sequential linear layers.

For transfer learning, the weights of the last linear layer of the base model are unfrozen, while the remaining linear layer weights are kept frozen. The knowledge learned from steel data is retained in the weights of these frozen layers.

The base model is first trained on the entire steel dataset. Then, in the data preprocessing step, the aluminum dataset is scaled using the same scaling parameters applied to the steel data. Finally, the model, with the last linear layer unfrozen, is trained on the aluminum data to develop aluminum models.

The transfer learning approach resulted in a slight improvement for both models in terms of MAE,  $R^2$  score, and NLL, as shown in Table 4.15, when compared to the DoE 3 method (Table 4.9). However, due to the prediction variance, the observed improvement cannot be considered significant.

Models	MAE	$R^2$	NLL
NN	11.88	0.82	
PNN	10.87	0.84	3.09

Table 4.15: DoE 5 - Performance metrics (Aluminum)

Figure 4.17 illustrates the prediction plots. It is observed that these plots closely resemble those from DoE 3 (Figure 4.13) and exhibit similar uncertainty in predictions.

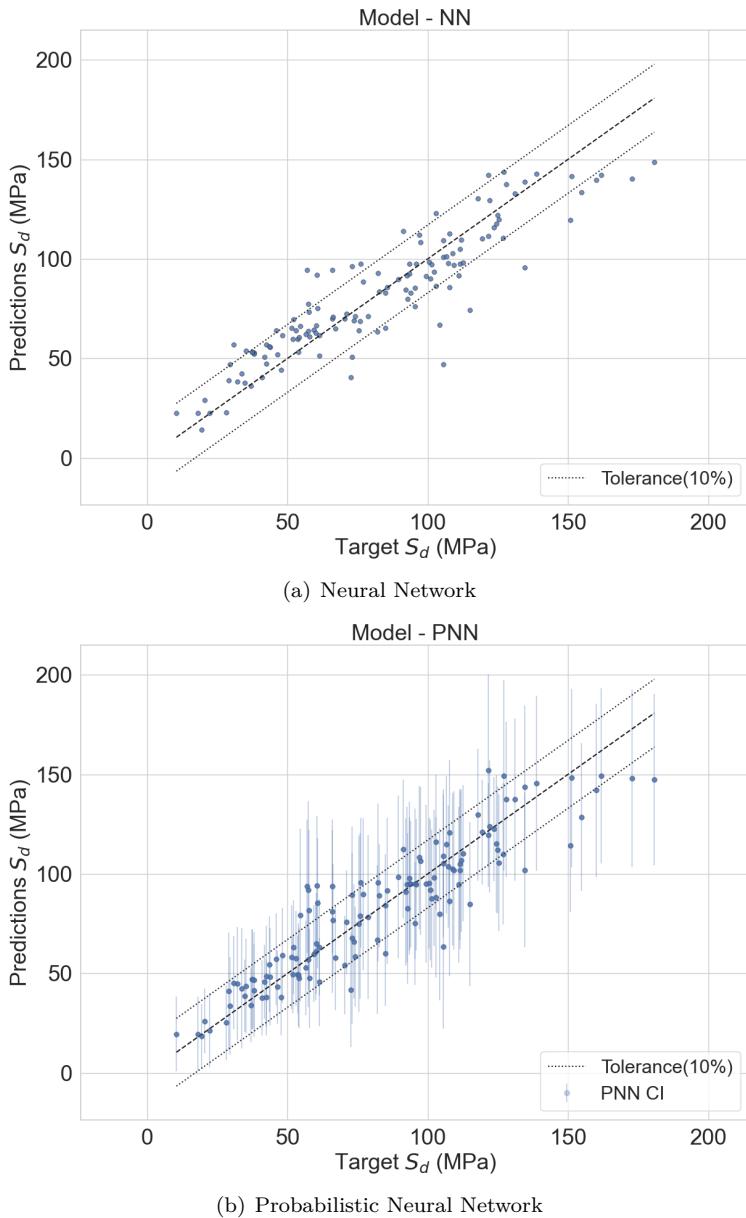


Figure 4.17: DoE 5 - Prediction plots for transfer learning (Aluminum)

## 4.4 Impact of data volume on model performance

Due to the limited availability of aluminum data, it is crucial to explore different approaches for improving prediction accuracy. One potential approach is increasing the dataset size, which theoretically should enhance model performance. Therefore, this section quantifies the probable improvement in accuracy with an increase in aluminum data.

This analysis is conducted by evaluating the performance metrics of the models using dataset sizes ranging from 30 to 130, with increments of 5 samples per step. For each dataset size, key performance indicators such as MAE and  $R^2$  score are computed. The trends of these performance indicators are then analyzed to estimate the potential accuracy improvement with further data expansion. This approach provides insights into the expected benefits of acquiring additional data and helps assess the extent to which increasing dataset size can enhance model performance.

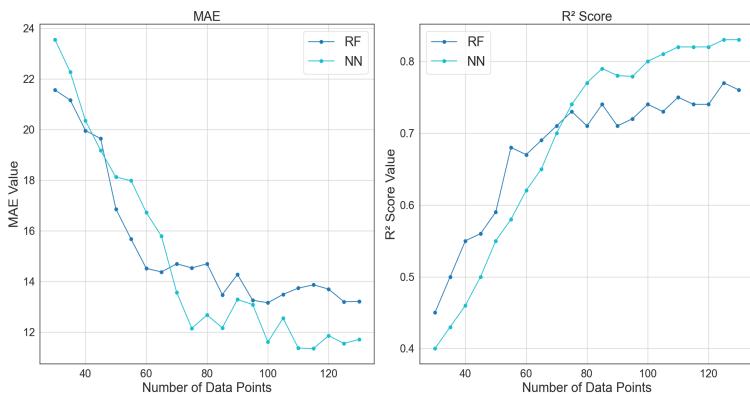


Figure 4.18: Performance metrics based on number of data points

From Figure 4.18, it is observed that after a dataset size of 80, both Random Forest (RF) and Neural Network (NN) reach a **plateau**, where

a further increase in data size does not significantly improve model performance. While additional data may still lead to gradual improvements, no substantial enhancement is evident beyond this point.

RF demonstrates the ability to achieve good results with a relatively small dataset. Similarly, NN also achieves strong performance with more than 80 data points, likely due to the lower complexity of fatigue limit prediction in this case.

As discussed, increasing the dataset size may not necessarily lead to a significant improvement in model accuracy. An alternative approach is to analyze regions with high prediction errors.

Since tensile strength ( $R_m$ ) and the stress concentration factor ( $K_t$ ) are key contributors to the fatigue behavior of metals, heat maps are generated to visualize the absolute prediction error and the corresponding number of data points within the feature space of  $R_m$  and  $K_t$ . These visualizations help establish correlations between these parameters and aid in identifying regions where additional experimental data could enhance model performance.

From Figure 4.19, it is evident that the highest prediction errors occur in regions with extremely high and low tensile strength, specifically below 250 MPa and above 600 MPa. This behavior is attributed to the limited number of data points available in these regions, as shown in Figure 4.20.

However, the model exhibits reliable predictions for data points corresponding to notched specimens with  $K_t > 3$ , despite their limited representation in the dataset. This can be attributed to their distinct mechanical properties compared to specimens with different  $K_t$  values.

Collecting additional data in the regions where  $R_m < 250$  MPa,  $R_m > 600$  MPa, and  $R_m < 300$  MPa with  $K_t > 3$ , where data is sparse, will further improve overall model accuracy.

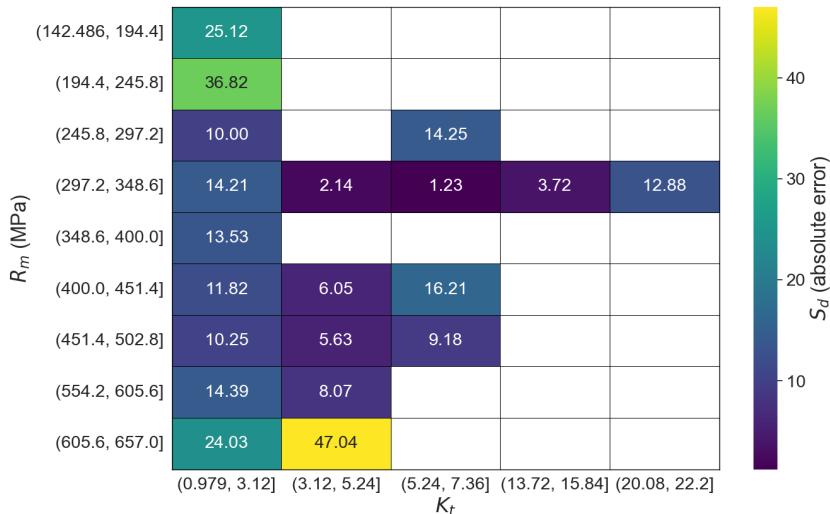


Figure 4.19: Fatigue limit ( $S_d$ ) absolute error in feature space  $R_m$ ,  $K_t$

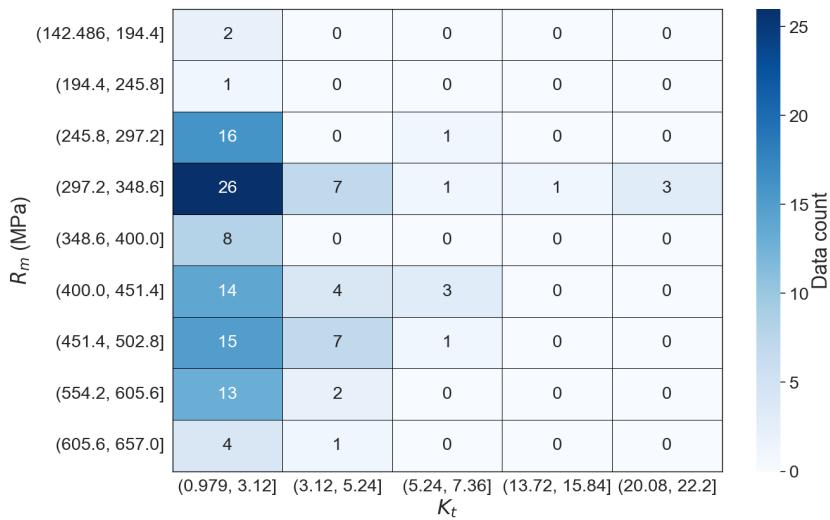


Figure 4.20: Data count in feature space  $R_m$ ,  $K_t$



## 5 Discussion and Conclusion

The findings of this study highlight the potential of machine learning models in predicting the fatigue life of metallic materials. The successful application of both deterministic and probabilistic models demonstrates the effectiveness of data-driven approaches in addressing the limitations of conventional fatigue analysis methods, such as the FKM analytical approach (refer to 2.1.4).

This research differentiates itself from previous studies in several ways. Previous research mainly focused on steel alloys [4, 10, 8] and relied on the relatively less diverse NIMS dataset [19]. In contrast, this study extends data-driven modeling for aluminum fatigue life prediction, thus demonstrating the advantages of machine learning over conventional fatigue analysis methods. Furthermore, this research broadens the steel database by incorporating greater data variability and increasing the number of data points for notched samples from sources such as JSMS [2] and DaBef [12].

The thesis workflow begins with understanding the concept of fatigue and the factors influencing it (refer to Chapter 2). Various approaches exist for fatigue parameter prediction, with laboratory testing and the FEA method posing significant drawbacks in terms of time and resource consumption. Since fatigue parameter estimation is a critical design factor that must be considered early in the design cycle, a quick estimation method is essential to reduce overall development time. The widely used FKM approach provides a simple analytical equations for fatigue limit

estimation. However, this method suffers from low prediction accuracy due to its reliance on simplified approximated equations. To overcome these challenges, data-driven approaches were explored, offering a balance between accuracy and computational efficiency.

For effective data-driven modeling, a high-quality dataset is necessary. Historical laboratory testing data was collected from multiple sources, and data preprocessing techniques were employed to transform the data into a usable format for machine learning models. Chapter 3 illustrates the relationships between tensile strength, stress concentration factor, stress ratio, load type, and stress gradient influence on fatigue limit. The correlation matrix revealed that tensile strength has a correlation value of 0.24 with fatigue limit (refer to Figure 3.5), consistent with the FKM analytical approach (Figure 2.10). Many studies estimate fatigue limits solely based on tensile stress, indicating its major contribution to fatigue life prediction. Other parameters play a smaller but still significant role.

## 5.1 Optimal ML models

The results from DoE 1 indicate that complex ML models significantly outperformed the FKM analytical approach, achieving a 50% reduction in MAE. While the FKM analytical method yielded an MAE of 22.53 MPa, complex ML models achieved MAE values in the range of 11.76 - 13.22 MPa (Table 4.5). However, the linear regression model performed similarly to the FKM analytical method, as both rely on simple analytical equations for fatigue limit estimation.

Figure 5.1 illustrates the performance of deterministic and probabilistic ML models across all DoE configurations for fatigue limit prediction in aluminum. Among deterministic models, Random Forest (RF) and Neural Networks (NN) exhibited superior performance with the lowest

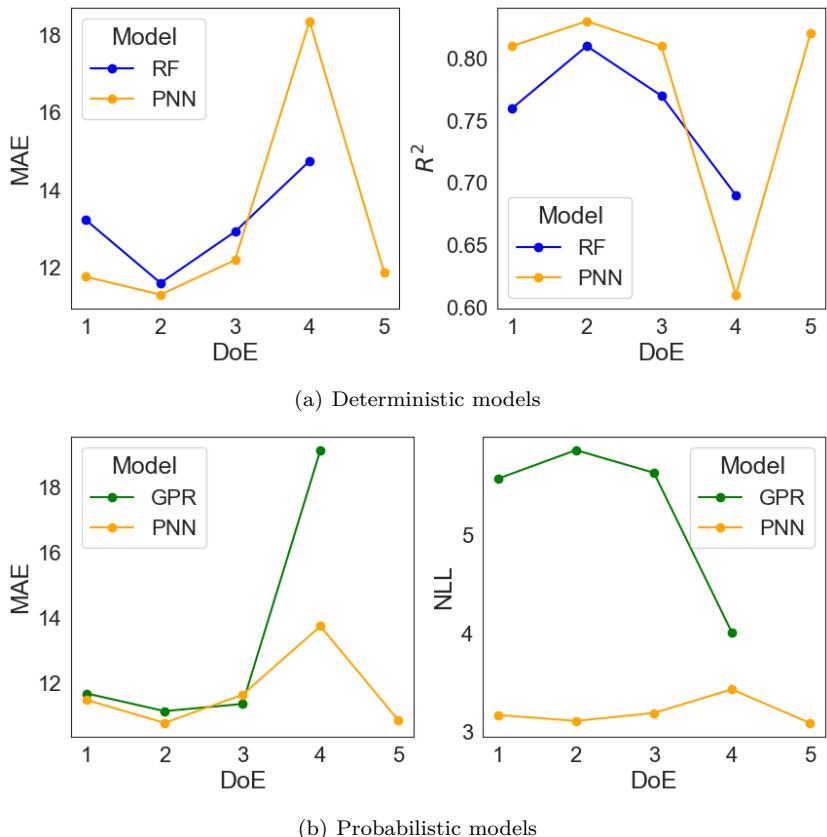


Figure 5.1: DoE comparison for aluminum models

MAE values. Within the probabilistic models, Gaussian Process Regression (GPR) and Probabilistic Neural Networks (PNN) achieved the best results, demonstrating the lowest NLL and MAE values (refer to 4.3). Consequently, the figure highlights only the performance of these models.

To determine the most effective model within each category, **NN consistently outperformed** RF across all DoE configurations, achieving lower MAE and higher  $R^2$  scores, as represented by the orange line in Figure 5.1. Similarly, **PNN outperformed** other probabilistic models, exhibiting significantly lower NLL values and slightly better MAE, as indicated by the orange line in Figure 5.1, which remains below other lines across all DoE configurations.

For the expanded steel dataset, the results indicate an MAE range of 30–35 MPa, with RF emerging as the top-performing deterministic model and PNN as the best probabilistic model (refer to 4.3). These findings align with previous studies, where RF and deep ensembles were identified as the best-performing models, achieving an average MAE of approximately 35 MPa using a similar feature set [8].

## 5.2 Optimal feature set

Various methodologies were explored to determine the best training strategy and feature set for predicting fatigue limits. This involved conducting a series of experiments (refer to Table 4.4). The results from the initial three experiments indicated that DoE 2 (Figure 5.1), utilizing feature set 2 (Table 4.3), yielded the lowest MAE error across all models. Notably, non-complex models like linear regression and k-nearest neighbors significantly improved by 50% due to the inclusion of the additional feature  $S_{d,FKM}$ , which increased model linearity. In contrast, complex

models such as RF, NN, GPR, and PNN showed minimal improvement, indicating the limited contribution of  $S_{d,FKM}$ .

Furthermore, DoE 1 and DoE 3 produced similar MAE error ranges, suggesting that  $G_0$  is not significantly relevant from a data-driven perspective. However, from an engineering standpoint, it contributes to fatigue limit estimation. In this study,  $G_0$  was approximated using the FKM method (refer to 2.1.3). Ideally,  $G_0$  should be derived using the FEA method, but due to computational and time constraints, FEA-derived  $G_0$  values were unavailable. Therefore, further investigation into the influence of  $G_0$  in data-driven modeling is necessary.

Based on the results of DoE 1, DoE 2, and DoE 3, it can be concluded that **feature set 3** (Table 4.3) is the optimal choice as it solely relies on experimental features, eliminating the need for analytical features  $G_0$  and  $S_{d,FKM}$ , thereby reducing computational efforts. However, from a product development perspective, **feature set 2** is preferable since it includes  $S_{d,FKM}$ , allowing the model to predict more accurately even for unseen data. For instance, the aluminum dataset lacks torsion load data. In such cases, a pure black-box model would yield high errors, whereas a grey-box model incorporating  $S_{d,FKM}$  enhances robustness in practical applications.

### 5.3 Optimal training method

Further, the results from DoE 3, DoE 4, and DoE 5 were analyzed to determine the optimal training method. These methods were evaluated using the same feature set, consisting only of experimental features (Table 4.3). From Figure 5.1, it is evident that DoE 4 yielded the worst performance among all configurations, as the predictive accuracy of all models deteriorated significantly. This approach involved combining

both steel and aluminum data into a single model with an additional categorical feature indicating the material type.

In contrast, DoE 5, which employs a transfer learning strategy, showed a slight improvement over DoE 3 for NN and PNN models. However, this improvement is marginal and cannot be considered significant due to the inherent variance in neural network predictions under identical model and data split seedings.

Therefore, DoE 3, which involves **training separate models** for steel and aluminum, is identified as the most effective training method.

Overall, this thesis successfully determined the optimal methodology for data-driven model development in predicting fatigue parameters across different metal classes. The best model performance was achieved when **separate models** were trained for each metal class using their respective datasets. Additionally, the results indicate that utilizing **only experimental features** is sufficient to develop a black-box model, eliminating the need for computationally expensive analytical or simulation-derived features.

This research extends previous studies [4, 10, 8] on data-driven fatigue life prediction of metallic materials by demonstrating robust methodologies for developing ML models specifically for aluminum.

## 5.4 Future scope

The success of previous data-driven modeling for fatigue life prediction in steel, along with the findings of this thesis for aluminum, paves the way for further research on other metals such as copper, titanium, and beyond.

As discussed earlier, the influence of the stress gradient ( $G_0$ ) on fatigue limit  $S_d$  prediction from a data-driven perspective remains an open question. Future studies should further investigate this aspect using FEA-simulated  $G_0$  values rather than analytically approximated ones.

Additionally, due to the lack of process-related features in the aluminum dataset, such as heat treatment, only engineering features were considered. However, prior research on steel [9] demonstrated that incorporating heat treatment data reduced the mean absolute error (MAE) by 12% in ML models. Therefore, exploring the impact of heat treatment features on aluminum fatigue life prediction is a worthwhile direction for future research.

As discussed in Chapter 4.4, the relationship between increasing dataset size and model performance improvement remains uncertain. However, given the limited data currently available, it would be valuable to investigate the model's performance with an expanded dataset. In particular, collecting data from higher-error regions could enhance the overall predictive accuracy.



# A Results of other fatigue parameters

This section presents the results for additional fatigue parameters, including the knee point ( $N_e$ ), the slope in the finite region ( $K$ ), and the scatter in the infinite region ( $T_s$ ). These results are obtained using the best methodology identified for predicting the fatigue limit ( $S_d$ ). Due to the varying scales of these fatigue parameters, direct physical interpretation becomes challenging. Therefore, to enhance intuition and facilitate comparison, all graphs and results presented in this section are normalized to the [0,1] range.

Fatigue parameters	Models	MAE	R <sup>2</sup>	NLL
$K$	RF	0.12	0.39	
	RFCI	0.12	0.40	3.09
$N_e$	RF	0.15	0.32	
	RFCI	0.14	0.32	4.04
$T_s$	RF	0.18	0.09	
	RFCI	0.18	0.08	4.75

Table A.1: Aluminum fatigue parameters performance metrics

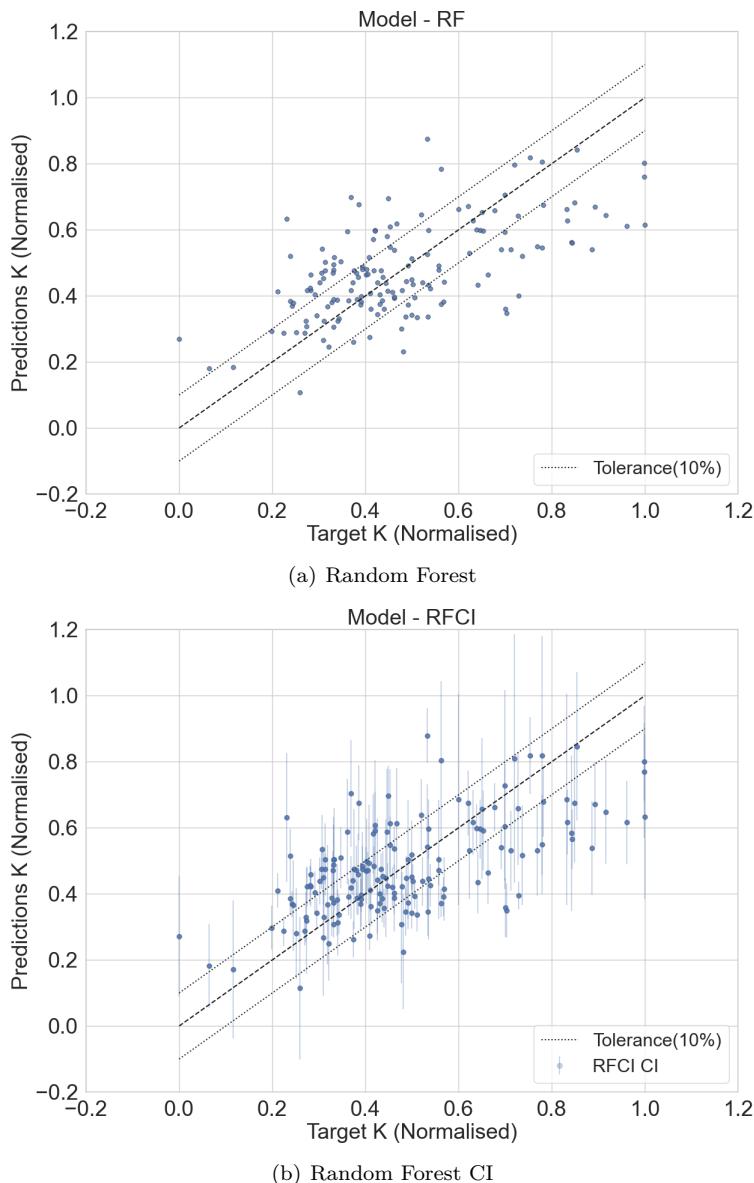
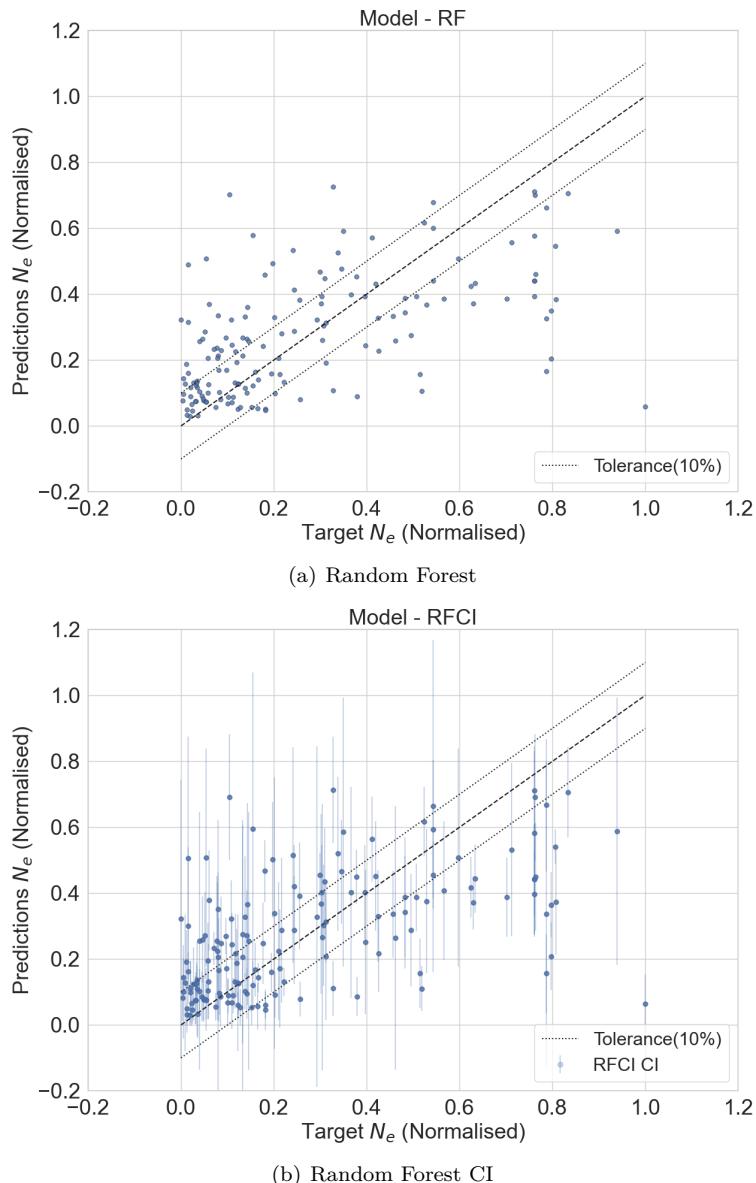


Figure A.1: Slope in finite region (K) prediction plots (Aluminum)

Figure A.2: Knee point ( $N_e$ ) prediction plots (Aluminum)

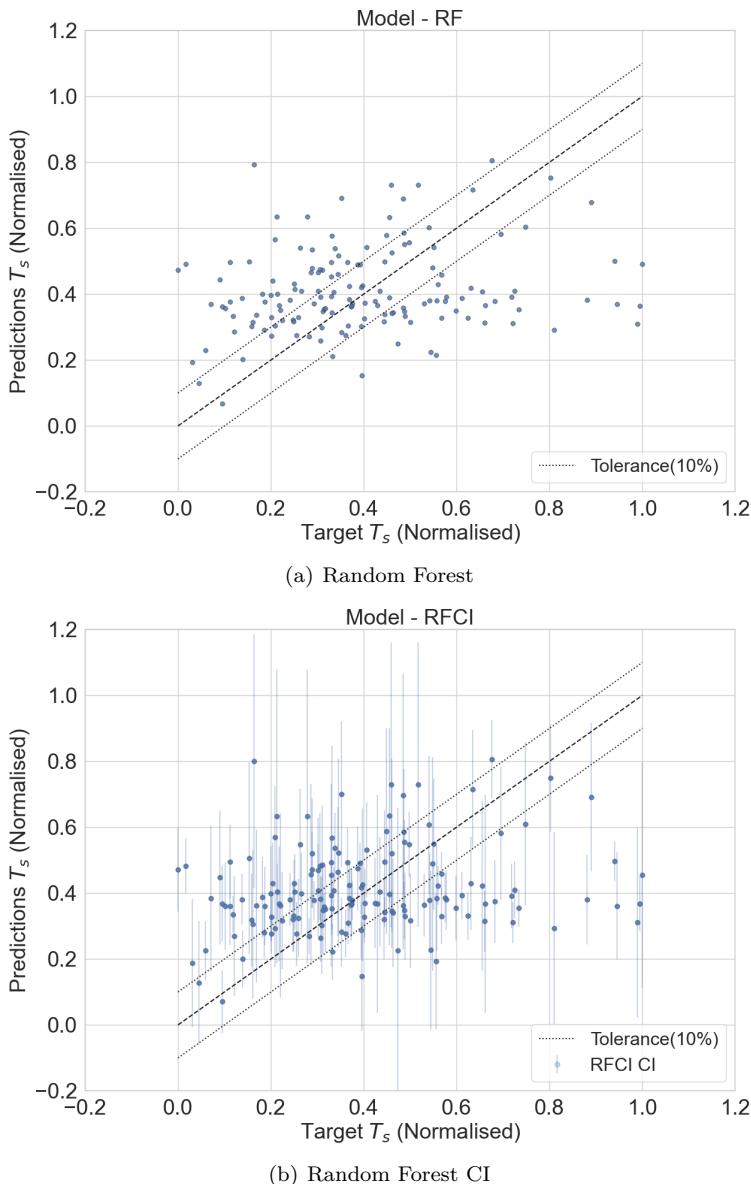


Figure A.3: Scatter in infinite region ( $T_s$ ) prediction plots (Aluminum)

## B Hyper parameters of machine learning models

Hyperparameter	Value
Number of estimators	500
Maximum depth	50

Table B.1: Hyperparameters - Random Forest.

Hyperparameter	Value
Number of neighbors	5

Table B.2: Hyperparameters - K Nearest Neighbor

Hyperparameter	Value
Constant	5
Epsilon	0.03

Table B.3: Hyperparameters - Support Vector Regressor

---

Hyperparameter	Value
Hidden size	[128, 256, 128]
Learning Rate	0.001
Optimizer	Adam
Activation Function	Sigmoid
Dropout Rate	0.4
Number of Epochs	2000
Patience	200 epochs
Early stopping	True
Batch normalization	True

Table B.4: Hyperparameters - Neural Networks

---

Hyperparameter	Value
<b>Kernel Structure</b>	$C \times (RBF + DotProduct)$
Constant Kernel ( $C$ )	$C(1.0, (10^{-3}, 10^3))$
RBF Kernel Length Scale	$\mathbf{1}$ (vector of ones)
RBF Length Scale Bounds	$(10^{-5}, 10^5)$
Dot Product Kernel $\sigma_0$	1
Dot Product $\sigma_0$ Bounds	$(10^{-5}, 10^5)$
Noise	$10^{-2}$

Table B.5: Hyperparameters - Gaussian Process Regressor

# Bibliography

- [1] SINOWON: *Fatigue testing machine ST-1180.* [https://www.sinowon.com/dynamic-static-fatigue-testing-machine-st-1180.](https://www.sinowon.com/dynamic-static-fatigue-testing-machine-st-1180) – [Photograph]
- [2] *Databook on fatigue strength of metallic materials.* Elsevier Science B.V, 1996
- [3] WAYKEN Rapid Manfacturing: *Metal strength.* <https://waykenrm.com/blogs/everything-you-need-to-know-about-metal-strength/>. Version: 2021. – [Photograph]
- [4] AGRAWAL, Ankit: Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. In: *Integrating materials and manufacturing innovation* 3 (2014), S. 90–108
- [5] AGRAWAL, Ankit: An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. In: *International Journal of Fatigue* 113 (2018), S. 389–400
- [6] BATHIAS, Claude: *Fatigue Limit in Metals.* ISTE Ltd 2014, 2014
- [7] C. GAO, J.C. P. M.Q. Yang Y. M.Q. Yang: Abnormal relation between tensile and fatigue strengths for a high-strength low-alloy steel. In: *Materials Science and Engineering* 832 (2022)

- [8] CHRISTIAN FRIE, Chris E. Anton Kolyshkin K. Anton Kolyshkin: Analysis of data-driven models for predicting fatigue strength of steel components with uncertainty quantification. In: *Fatigue Fracture of Engineering Materials Structures* 47 (2024), Nr. 3, S. 1036–1052
- [9] CHRISTIAN FRIE, Chris E. Anton Kolyshkin K. Anton Kolyshkin: Exploration of materials fatigue influence factors using interpretable machine learning. In: *Fatigue Fracture of Engineering Materials Structures* 47 (2024), Nr. 9
- [10] DORINA WEICHERT, Sebastian Houben Gunar Ernis Stefan W. Alexander Kister K. Alexander Kister: Robustness in Fatigue Strength Estimation. (2022)
- [11] DR. ING. ROLAND RENNERT, Prof. Dr. Ing. Michael Vormwald Prof. Dr. Ing. Alfons Esderts Dr. Ing. Dieter S. Prof. Dr. Ing. Eckehard Kullig K. Prof. Dr. Ing. Eckehard Kullig: *ANALYTICAL STRENGTH ASSESSMENT OF COMPONENTS*. VDMA Verlag, 2012
- [12] F. ELLMER, K.-G. E. A. Esderts E. A. Esderts ; HINKELMANN, K.: Forschungskuratorium Maschinenbau: Datenbank und Auswertesystem Betriebsfestigkeit. In: *FKM-Vorhaben* (2011)
- [13] FILIPPINI, Mauro: Stress gradient calculations at notches. In: *International Journal of Fatigue* 22 (2000), S. 397–409
- [14] GIUSEPPE, Bonaccorso: *Machine learning algorithms*. Packt Publishing Ltd, 2018
- [15] J.C. PANG, Z.G. Wang Z.F. Z. S.X. Li L. S.X. Li: General relation between tensile strength and fatigue strength of metallic materials. In: *Materials Science and Engineering* 564 (2013), S. 331–341
- [16] J.R, Davis: Aluminum and Aluminum Alloys. In: *Alloying: Understanding the Basics* (2001), S. 351–416

- [17] RAFAEL G. MANTOVANI, Edesio Alcobaça Joaquin Vanschoren André C.P.L.F. de C. André L.D. Rossi R. André L.D. Rossi: A meta-learning recommender system for hyperparameter tuning: Predicting when tuning improves SVM classifiers. In: *Information Sciences* 501 (2019), S. 193–221
- [18] REITERMANOVA, Zuzana: Data splitting. In: *WDS* (2010), S. 31–36
- [19] Y. FURUYA, H. Hirukawa N. N. H. Nishikawa N. H. Nishikawa ; TAKEUCHI, E.: Catalogue of NIMS fatigue data sheets. In: *Science and Technology of Advanced Materials* 20 (2019), S. 1055–1072
- [20] YUKITAKA MURAKAI, Kentaro Wada Hisao M. Toshio Takagi T. Toshio Takagi: Essential structure of S-N curve: Prediction of fatigue life and fatigue limit of defective materials and nature of scatter. In: *International Journal of Fatigue* (2021), S. 106–138
- [21] YUNG-LI LEE, Richard Hathaway Mark B. Jwo Pan P. Jwo Pan: *Fatigue testing and analysis*. Elsevier, 2005



## Statutory declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Siegen, March 4, 2025



Chirag Angadi

---

— name of the student —