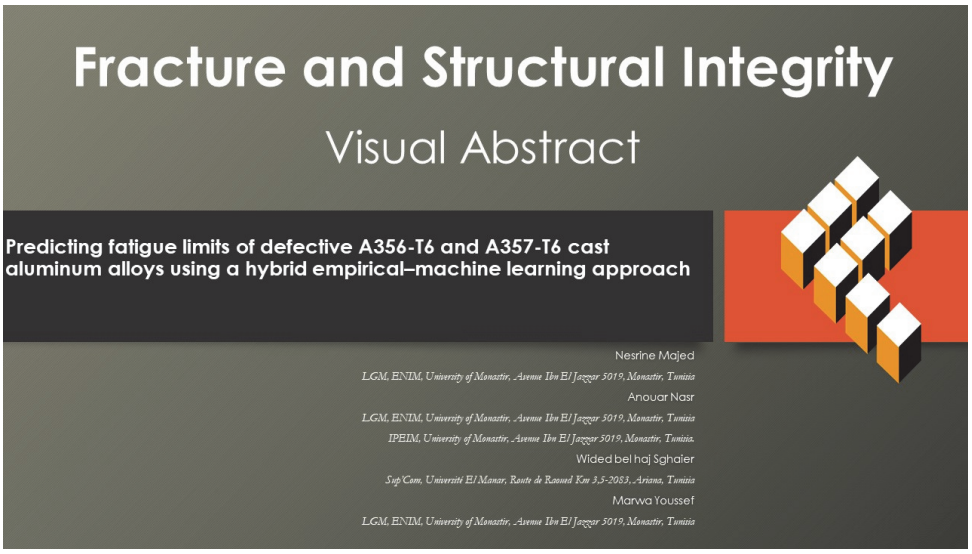
Predicting fatigue limits of defective A356-T6 and A357-T6 cast aluminum alloys using a hybrid empirical–machine learning approach

Nesrine Majed
LGM, ENIM, University of Monastir, Avenue Ibn El Jazzar 5019, Monastir, Tunisia
nesrine.majed@gmail.com, http://orcid.org/0009-0009-9157-8003

Anouar Nasr
LGM, ENIM, University of Monastir, Avenue Ibn El Jazzar 5019, Monastir, Tunisia
IPEIM, University of Monastir, Avenue Ibn El Jazzar 5019, Monastir, Tunisia.
anouar.nasr@hotmail.fr, http://orcid.org/0000-0002-2152-9910

Wided Bel Haj Sghaier
Sup'Com, Université El Manar, Route de Raoued Km 3,5-2083, Ariana, Tunisia
belhajsghaier.wided@gmail.com, http://orcid.org/0000-0002-5354-8841

Marwa Youssef
LGM, ENIM, University of Monastir, Avenue Ibn El Jazzar 5019, Monastir, Tunisia
marwayousseff@gmail.com, http://orcid.org/0000-0002-4902-2097



Citation: Majed, N., Nasr, A., Bel Haj Sghaier, W., Youssef, M., Hybrid empirical–machine learning approach for fatigue limit prediction of defective A356-T6 and A357-T6 cast aluminum alloys, Fracture and Structural Integrity, 76 (2026) 265-276.

Received: 07.10.2025
Accepted: 11.02.2026
Published: 18.03.2026
Issue: 04.2026

KEYWORDS. Fatigue, A357-T6, A356-T6, defect, Kitagawa, Machine learning.

INTRODUCTION

Due to its exceptional combination of outstanding corrosion resistance, high specific strength, good castability, and potential for improved CO_2 emission control through weight reduction, Al-Si-Mg cast alloys are widely used in automotive applications [1].

Hypoeutectic Al-Si cast alloys frequently solidify to form dendritic formations. The cooling rate has a significant influence on the final dendritic topologies during the solidification process [2]. The secondary dendritic arm spacing (SDAS) is a microstructural parameter frequently employed to measure certain dendritic structures [3]. SDAS is the parameter governing the fatigue limit for defect-free [4].

The casting defects, such as gas holes and shrinkage cavities, caused by the solidification process, have a significant impact on the fatigue behavior of cast aluminum alloys [5, 6]. Shrinkage cavity defect sizes could be characterized by $\sqrt{area}$ [7]. Murakami et al. [7] defined this parameter as equal to the square root of the defect surface on a plane perpendicular to the direction of maximum principal stress, to characterize the defect size. This parameter could be used to measure different defective morphologies. Karash [8] investigated the impact of stress ratio adjustments (R) on the fatigue fracture propagation rates for aluminum alloys. The findings showed that an increase in the positive stress ratio leads to an increase in the fatigue crack growth rate (reducing the cycles to failure), while a negative stress ratio (R=-1) causes a reduction in the rate at which fatigue cracks grow.

Since fatigue testing requires multiple load cycles and extended periods in order to ensure structural safety, it is both expensive and time-consuming [9]. When combined with physics-based, empirical or semi-empirical models, current developments in machine learning significantly lower the volume of costly and time-consuming experimentation [10]

To predict the fatigue life of seven aluminum alloy series from tiny experimental datasets, Lian et al. [11] introduced a knowledge-based machine learning framework that combines empirical formulas with data-driven modelling. A gradient boosting regression model with features from the suggested estimation and guesswork techniques produced a mean relative error of 140%.

Utpat et al. [12] estimated the fatigue strength of cast aluminum alloy using machine learning methods. The models were developed using a dataset of 39 cast aluminum alloys. For model creation, four machine learning techniques were chosen: Random Forest (RF), Support Vector Machine (SVR), Artificial Neural Network (ANN), and Linear Regression (LR).

Che and Peng [13] proposed a SMA-SVR hybrid method. This combined model used the Slime Mould Algorithm (SMA) with support vector regression (SVR) for predicting two mechanical parameters tensile strength and 0.2% proof stress of low-alloy steel.

To reduce the dependency on expensive and time-consuming experimental testing (S-N technique), the authors Bhat et al [14] used machine learning (ML) to estimate the fatigue strength (endurance limit) of cobalt-based alloys. This study demonstrates that ANN-based machine learning models can accurately predict fatigue strength, offering a suitable alternative to conventional techniques for predicting the performance of cobalt alloys under cyclic stress.

To address data scarcity and accelerate the characterization of fatigue in aluminum alloys, Esmaeili et al. [15] presented a hybrid framework that combines an empirical model with data-driven methodologies. It was found that neural networks and SVR have a better performance.

For cast aluminum alloys like A356-T6 and A357-T6, the primary challenge to predicting fatigue life is the need for expensive and time-consuming experimental testing to guarantee structural safety. Although knowledge-based or hybrid frameworks have been proposed in recent machine learning (ML) studies for reducing the dependency on extensive testing, these data-driven approaches frequently encounter a recurring problem: the lack of available experimental data. This study presents a knowledge-driven hybrid framework that links machine learning and conventional empirical models. The main innovation is the creation of a significant synthetic dataset (5,000 points) using a calibrated empirical equation as a surrogate model, ensuring the preservation of basic physical correlations between SDAS and defect size. This method clearly shows cross-alloy transfer by successfully validating models trained on A356-T6 against experimental data from the A357-T6 alloy, in contrast to earlier research that concentrated on single-alloy estimation. The main distinction from earlier approaches is the method used to guarantee adequate data quality and variability: an empirical equation is used to create a large synthetic dataset in order to address the lack of available experimental data. The basic physical relationships between fatigue limit, through SDAS, and defect size ($\sqrt{area}$) are maintained this basis. By applying the SVR model on A356-T6 data (augmented with synthetic points) and then successfully testing its predictive ability on the related cast aluminum alloy, A357-T6, this study explicitly demonstrates the model's global performance, whereas prior works concentrated on estimation. To create the Kitagawa diagram of A356-T6 cast aluminum alloys, the objective of this work is to develop and train machine learning

models, including Support Vector Regression (SVR), Random Forest (RF), and Gaussian Process Regression (GPR). An extensive synthetic dataset is generated using an empirical equation, because experimental data is frequently limited. The predictive power and generalisation performance of the trained models are then evaluated on a cast aluminum alloy that shares many of the same properties as A357-T6. Tab. 1 summarizes the most relevant recent studies focusing on the fatigue life prediction of some metals using ML techniques.

References	Target material, Key feature (Dataset)	Models investigated	Observations
Chen *et al* [16]	Stainless steel, aluminum alloy, titanium alloy, magnesium alloy, alloy steel, copper alloy, and nickel alloy (Source: 36 articles, Size 1167)	CNN (Convolutional Neural Network) +FCL (fully connected layers) LSTM (Long Short-Term Memory Networks) +FCL GRU (Gate Recurrent Unit) +FCL SVM and RF	• Best R^2 =0.915431 • GRU+FCL • Perform better to predict fatigue life, Average error 5,6%
Hills *et al* [17]	Ti–6Al–4V additively manufactured (AM) parts. Dataset compiled from 55 studies, 143 samples, 23 input features (surface roughness, post-treatment, porosity, stress ratio, HIP, etc.)	Gradient Boosting Decision Tree (GBDT) Feature selection for sensitivity analysis	•R^2 = 0.80 using all selected sensitive features. •R^2 = 0.69 when using only practically measurable features (surface roughness, post-treatment, etc.). • Confirms strong influence of surface roughness, porosity, HIP, and stress ratio on fatigue strength.
Shin *et al* [18]	Metallic materials under low-cycle fatigue (LCF), using stress–strain time-series from cyclic tests (single-cycle data)	Deep learning: LSTM + CNN	• R^2 = 0.99 for life prediction, model outperforms baseline •LSTM and CNN models
Zhang *et al* [19]	2024-T3 clad aluminum alloy, uniaxial fatigue specimens; dataset includes surface roughness, stress concentration factor, mean stress,	Hybrid: improved physical SWT (Smith Watson Topper) criterion + ML XGBoost (extreme gradient Boosting), Random Forest + parameter optimization	• R^2 = 0.96 for fatigue life prediction using the ensemble "blending" model, outperforming the individual GA (Genetic Algorithm) -HL (Hyperparameter Learning) - XGBoost, GA-RF, and Deep Random Forest models.
Li *et al* [20]	Aluminum alloys dataset of 54 S–N curves across 7 different aluminum alloys	Deep operator learning with transformer-based encoder + domain-informed features + ML baseline comparisons	$R^2 \approx 0.9515$, mean absolute error ≈ 0.2080, mean relative error ≈ 0.5077 significantly better than standard ML/DL baselines

Table1: Summary of studies on fatigue life prediction using ML techniques.

MATERIALS AND DATASET

In the present work, an empirical equation is used to conclude a synthetic dataset representative of the fatigue response of defective cast aluminum alloys. This procedure was adopted to overcome the scarcity of available experimental data and in order to ensure sufficient variability in the input parameters. The generated dataset preserves the fundamental physical correlations between defect size, stress amplitude, and fatigue limit. Such a data foundation enables the application of machine learning algorithms for predictive modeling and uncertainty quantification.

Materials

Tab. 2 represents the chemical composition of the two-cast aluminum alloys A356-T6 and A357-T6 [15].

Alloy type	Si	Mg	Mn	Zn	Cu	Cr	Ti	Fe
A 356	7	0.35	0.1	0.1	0.2	0	0.2	0.2
A 357	7	0.57	0.05	0.05	0.1	0	0.12	0.12

Table 2: Chemical composition of the cast aluminum alloys [15].

Data collection

The dataset includes the two tables of fatigue limits for two cast aluminum alloys, which were collected from the previous literature. The first dataset, which is represented in Tab. 3 [4], consists of 16 fatigue limits of the cast aluminum alloy A356-T6 specimens characterized by an SDAS value and containing a spherical surface defect measured by the Murakami [21] parameter. The experimental dataset was divided into calibration and test subsets, comprising 70% and 30% of the data, respectively. The distinction between the two subsets is explicitly indicated in Tab. 3.

SDAS [μm]	$\sqrt{area}$ [μm]	Fatigue limit [MPa]	R	Dataset
38	50	90	-1	
57	500	75	-1	
39,5	750	65	-1	
38	400	85	-1	
39	900	58	-1	calibration
39	600	80	-1	
39,5	380	88	-1	
38	700	75	-1	
58	700	70	-1	
58	50	91	-1	
39,5	520	82	-1	
38	600	80	-1	
39,5	0	90	-1	
58	100	90	-1	test
38	750	65	-1	
58	500	75	-1	

Table 3: Fatigue limits under tensile loading, R= -1, for spherical defects A356-T6 [4]

The second dataset is presented in Tab. 4 [22], which comprises 7 fatigue limits of cast aluminum alloy A357-T6, with an average SDAS of 38 mm for various spherical defect sizes.

SDAS [μm]	Defect size $\sqrt{area}$ [μm]	Fatigue limit [MPa]	Load ratio R
38	50	90	
38	50	100	
38	200	91	
38	400	90	-1
38	400	80	
38	600	80	
38	880	70	

Table 4: Fatigue limits under tensile loading, R=-1, for spherical defects A357-T6 [22]

Dataset generation

Because there was not enough experimental data available, Ben Houria's empirical equation was used to create a larger dataset [4]. The term y represents a constant offset introduced to account for the systematic mean deviation between the experimental fatigue strength values and the polynomial component of the model. In this study, 70% of the available experimental data were used to estimate the model coefficients, including the constant offset y, via the least-squares method, while the remaining 30% were reserved for independent validation. An empirical equation previously proposed in Reference [4] was employed to describe the fatigue limit of the A356-T6 alloy as a function of SDAS and $\sqrt{area}$:

$$\sigma = a * SDAS^2 + b * \sqrt{area}^2 + c * SDAS * \sqrt{area} + y \quad (1)$$

where a, b, c, and y denote the polynomial coefficients obtained by fitting the experimental data [4]. As a result, during the creation of the synthetic data, y is handled as a deterministic constant and maintained constant across all data points.

$a = -2 * 10^{-3}$, $b = -4.16 * 10^{-5}$, $c = -9.63 * 10^{-6}$ and y = 95.92 MPa

Using the empirical equation, a dataset of 5000 points was created for this study, considering $\sqrt{area}$ values between 0 and 900 μm and SDAS values between 25 and 80 μm. The interval [25–80 μm] used in our synthetic dataset is chosen to cover the experimentally observed SDAS range in A357-T6 alloys. Wang et al. [23] measured SDAS over a wide range of cooling rates and obtained SDAS values typically between ≈20 μm and more than 80 μm, depending on the solidification conditions and mold type. Fig. 1 shows the workflow scheme of the generation process of 5000 synthetic fatigue limit.

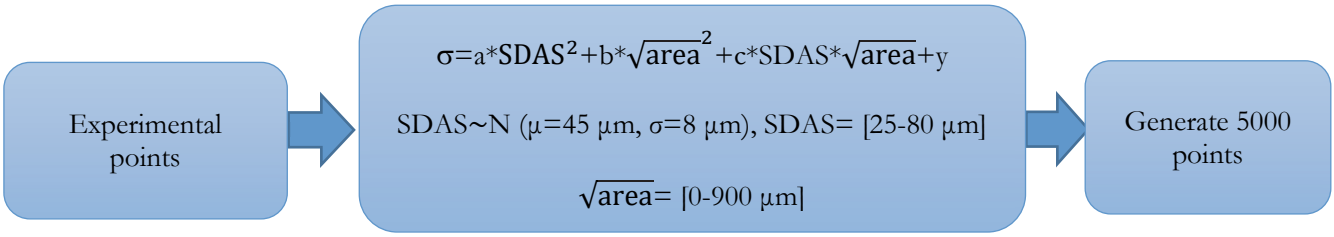


Figure 1: The workflow scheme of the generation process.

The secondary dendrite arm spacing (SDAS) was assumed to follow a normal (Gaussian) distribution, characterized by a mean value of 45 μm and a standard deviation of 8 μm, in accordance with experimental observations reported in the literature [22, 24].

This approach allows the machine learning model to be used as a surrogate model to reproduce the empirical fatigue relationship under controlled conditions.

The generation process follows a sequential seven-step approach to ensure reproducibility and physical relevance:

- Initialization: A random seed (rng (42)) is fixed to ensure the reproducibility of the generated results.
- Model Parameterization: The coefficients for the empirical equation are defined based on established literature values: a=−2.00×10−3, b=−4.16×10−5, c=−9.63×10−6, and y=95.92 MPa.

- SDAS Generation: SDAS values follow a truncated normal distribution N $(45, 8^2)$ µm, to respect experimentally observed physical limits. A rejection sampling approach is employed until N=5000 points are reached.
- $\sqrt{area}$ Generation: $\sqrt{area}$ follows a uniform distribution U (0,900) µm, covering the full range of observable defect sizes in cast aluminum alloys.
- Stress Calculation: The fatigue limit (σ) is calculated for each (SDAS, $\sqrt{area}$) pair according to the parametric Eqn. (1)
- Load Ratio Assignment: The load ratio (R) is fixed at -1 for the entire dataset, modeling symmetric alternating loading typical of constant-amplitude fatigue testing.
- Data Export: The final dataset, comprising 5000 data points, is structured in a MATLAB table and exported to Excel format for use in the machine learning algorithms.

This methodology ensures that the generated basis preserves the fundamental physical correlations between fatigue limit, SDAS, and defect size.

MACHINE LEARNING MODELS

Fig. 2 shows the schematic workflow of model creation used in this study to estimate Kitagawa diagrams of aluminum cast alloys. The next stage is to separate the data into training and testing subsets after creating two datasets and choosing pertinent features.

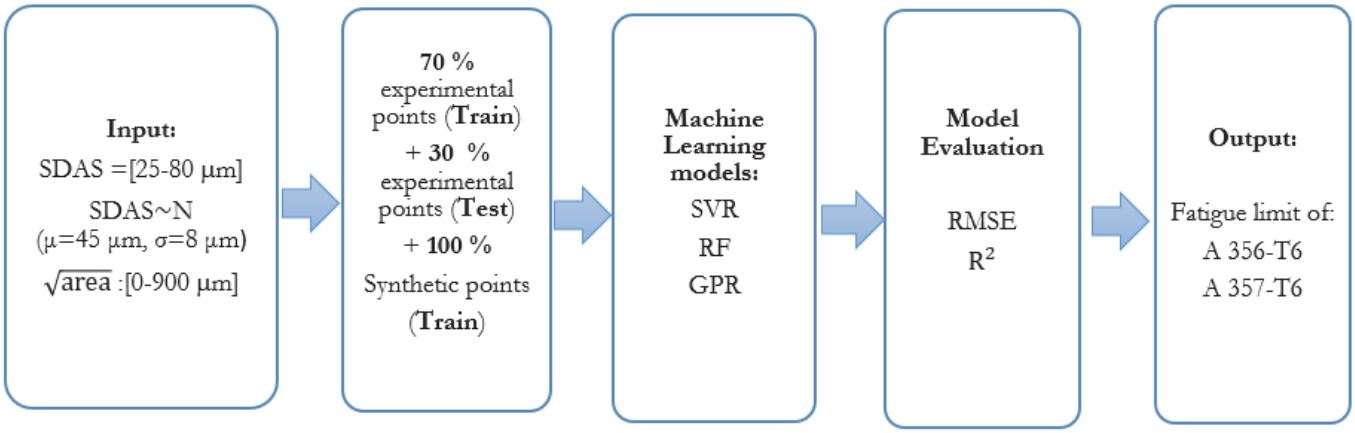


Figure 2: Schematic workflow of the model development for aluminum cast alloys.

The testing data, which are not exposed during training, are saved to validate the created models, whereas the training data are used to train the models. Experimental and synthetic data were combined to create the dataset that was used to train and evaluate the machine learning models. Initially, a training and a test subset were randomly selected from the available experimental database. We added 100% of the synthetic data produced by the empirical polynomial relationship between SDAS, √area, and fatigue strength to the training set, which comprised 70% of the experimental data. The remaining 30% of the experimental data made up the test set (independent validation set); no synthetic points were included. In this way, both experimental and augmented (synthetic) data are used to train the models, and only unobserved experimental fatigue data are used to evaluate their capacity for generalization.

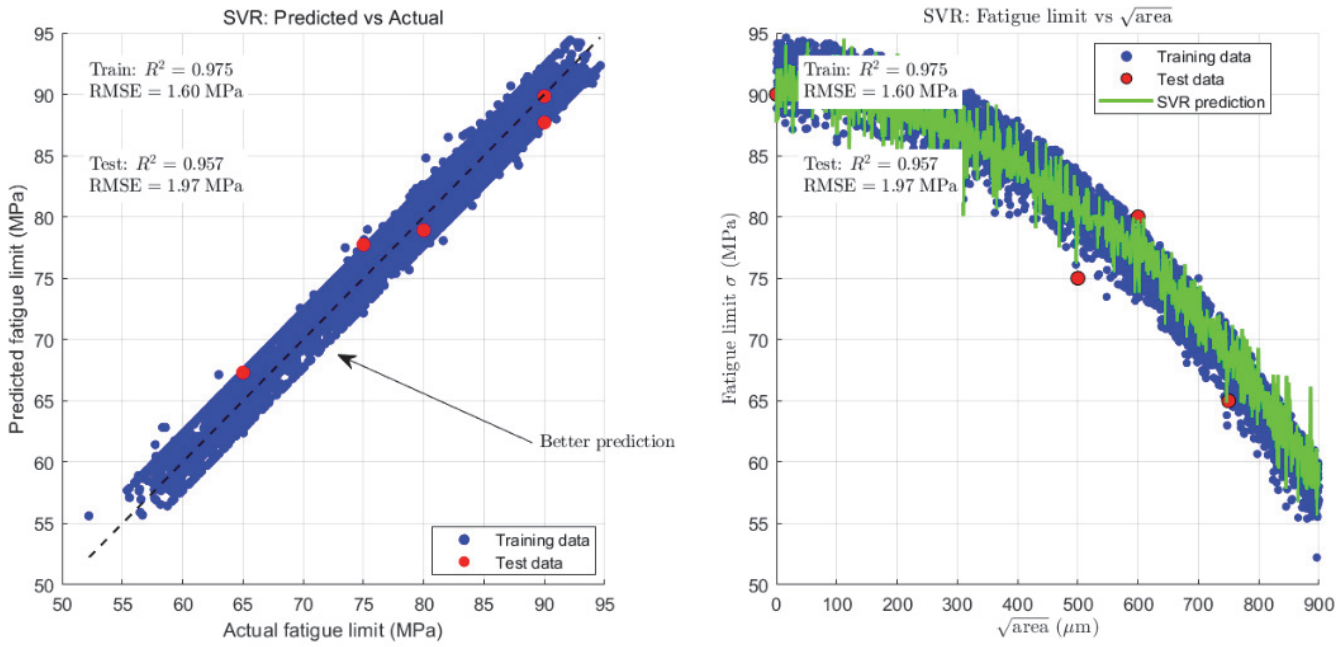
The foundation of any machine learning algorithm is the selection and optimization of an appropriate model. Generally, we should experiment with various algorithms and evaluate the results. This study assessed both traditional and cutting-edge machine learning techniques, such as Gaussian Process Regression (GPR), Random Forest (RF), and Support Vector Machine (SVM). Regression plots were used to evaluate the efficiency of the prediction models. The vertical axis in these charts (see Fig. 3a, Fig. 4a and Fig. 6a) displays the predicted values, while the horizontal axis represents the actual data values. The plot's sample points are all positioned according to both the predicted and actual values. All the sample points line up precisely along a straight line that runs down the plot's diagonal when the expected and actual values coincide.

Support Vector Regression Model for A356-T6

Vapnik et al. [25] proposed SVR in 1996. SVR aims at approximating the functional relationship between input vectors $x \in R^n$ and the output $y \in R$. The regression function is expressed as:

$$F(x)=\langle w, \Phi(x)\rangle + b \quad (2)$$

where $\phi(x)$ denotes a nonlinear mapping to a high-dimensional feature space, w is the weight vector, and b is the bias term [26].



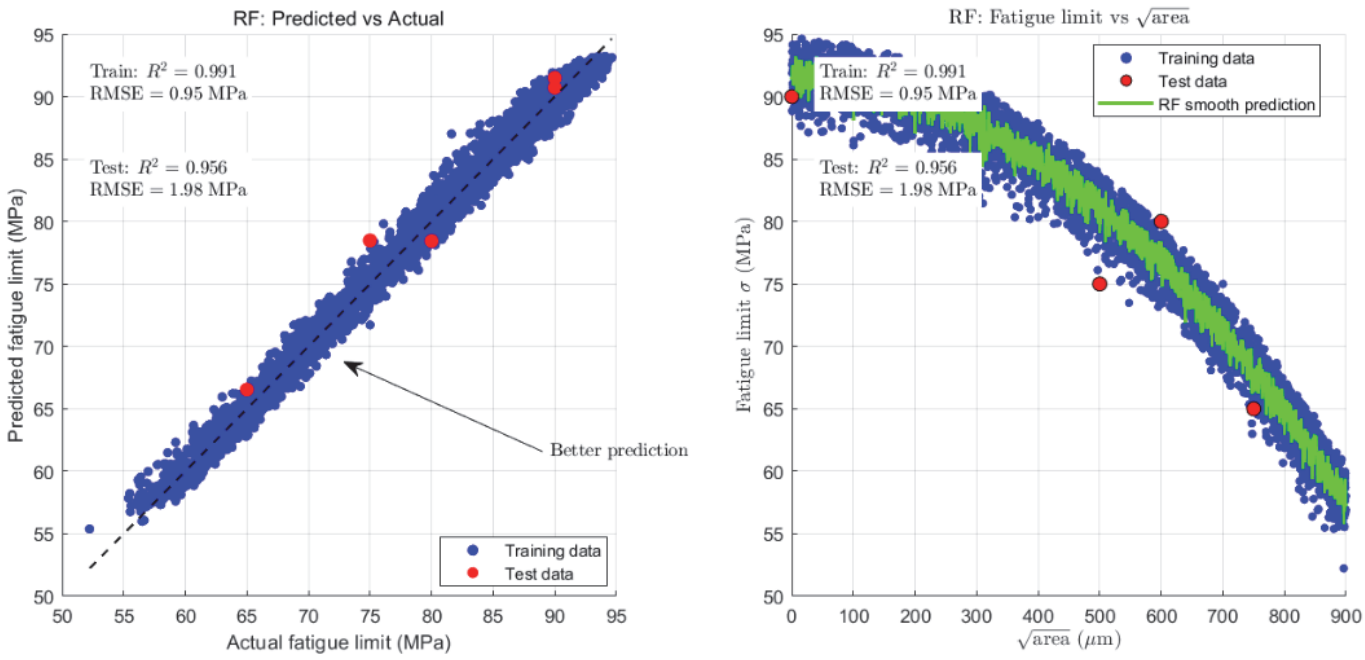
(a) (b)

Figure 3: (a) The regression plot of the SVR model for the A357-T6 cast aluminum alloy. (b) Kitagawa Diagram of A356-T6 under tension loading, R = -1, using the SVR model.

The Kitagawa diagram of the cast aluminum alloy A356-T6 under reversed tensile loading conditions (R = -1) is shown in Fig. 3b. The experimental test results are displayed as red symbols, and the support vector regression (SVR) model is plotted as the blue continuous curve. With a coefficient of determination on the test set of $R^2 = 0.957$, the SVR model exhibits an RMSE (Root Mean Square Error) is equal to 1.97 MPa. This high accuracy suggests that the SVR approach can accurately capture the fatigue strength transition between the defect-dominated regime (at large defect sizes) and the defect-insensitive regime (at small defect sizes). The SVR model's robust aspect in characterizing the nonlinear relationship between the stress amplitude and $\sqrt{area}$ is further demonstrated by the smooth green curve it provides, without overfitting fluctuations in the experimental points. When all factors are considered, the outcomes demonstrate that the SVR model is a trustworthy method for forecasting the Kitagawa behavior of A356-T6 alloy when casting defects are present.
According to the statistics, most sample points in the SVR model's regression plot are grouped near the least-squares.

Random Forest for A 356-T6

The machine learning model was constructed using a traditional nonparametric model called Random Forest (RF) [27].
The Kitagawa diagram for cast aluminum alloy A356-T6 under fully reversed tension (R = −1) is displayed in Fig. 4b. With test data shown in red and training data shown in blue, the Random Forest prediction (green curve) closely matches the overall trend between stress amplitude and $\sqrt{area}$, producing $R^2 = 0.956$ and RMSE = 1.98MPa.

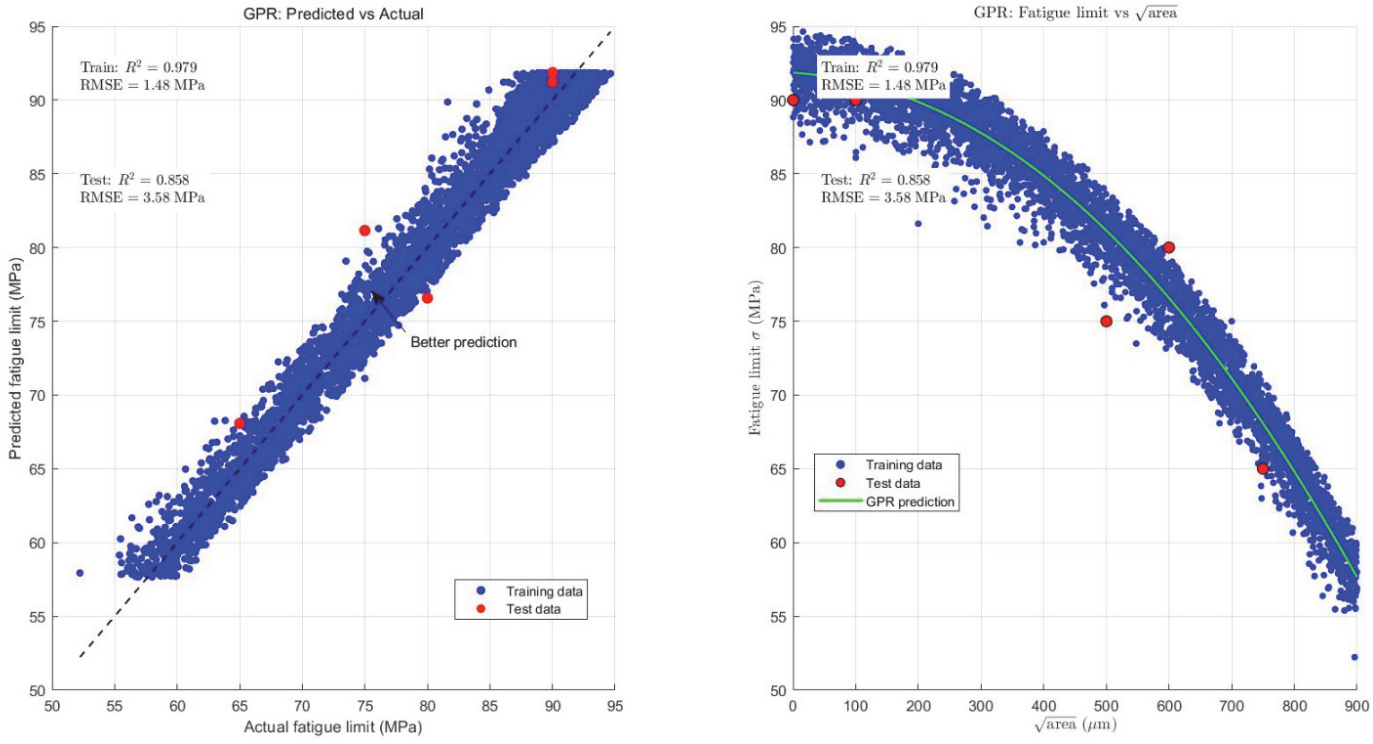


(a) (b)

Figure 4: (a) The regression plot of the RF model for the A356-T6 cast aluminum alloy, (b) Kitagawa Diagram of A356-T6 alloy under tension loading, R= -1 using the RF model.

Gaussian Process Regression for the A356-T6

A non-parametric Bayesian method called Gaussian Process Regression (GPR) [28] uses a multivariate Gaussian distribution over a set of data points to represent a set of random variables.



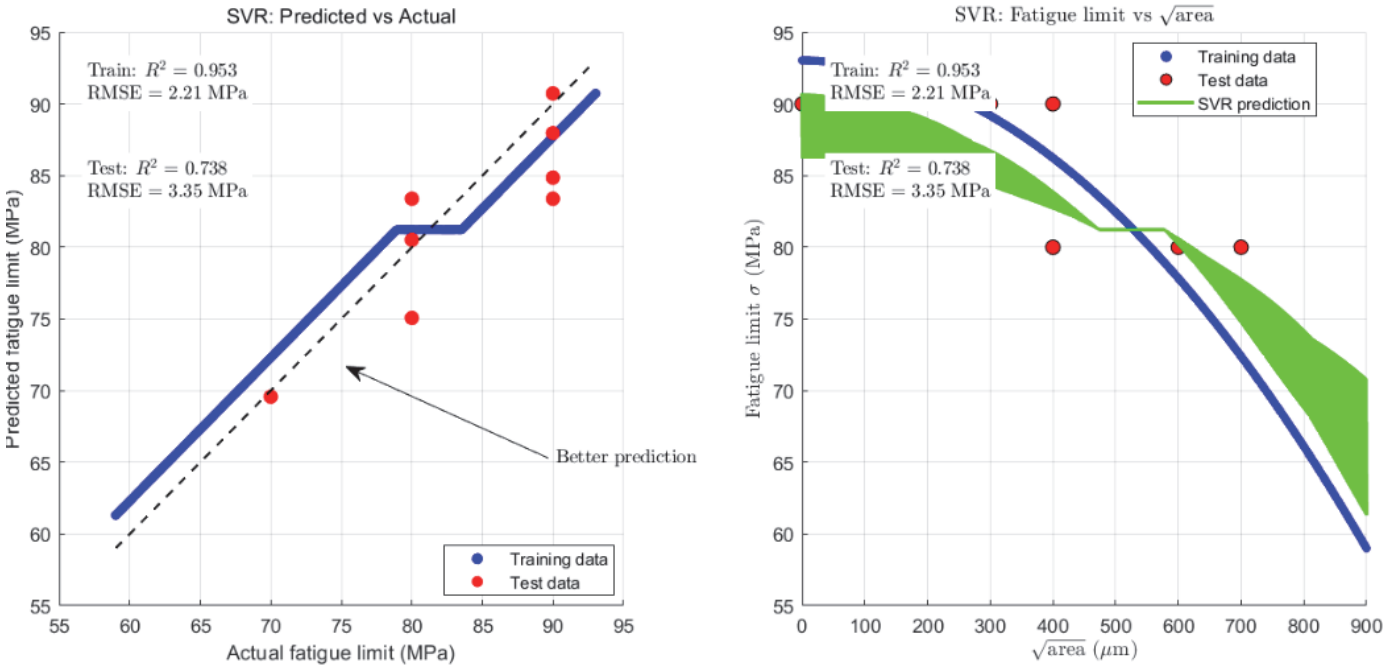
(a) (b)

Figure 5: (a) The regression plot of the GPR model for the A356-T6 cast aluminum alloy, (b) Kitagawa Diagram of the cast aluminum alloy A356-T6 under tension loading, R= -1, using the GPR model

Training data (blue) and test data (red) are well represented by the GPR model (green curve), which fits the Kitagawa diagram of A356-T6 very well, as seen in Fig. 5b. The robust aspect of the model in predicting the relationship between fatigue strength and defect size under fully reversed tensile loading (R = −1) is confirmed by the coefficient of determination (R^2 = 0.858) and RMSE = 3.58 MPa.

Support Vector Regression Model for A357-T6

We trained an SVR model for a fixed value of SDAS = 38 µm for the cast aluminum alloy A356-T6 and subsequently tested it on A357-T6 for the same SDAS value (Tab. 4).



(a) (b)

Figure 6: (a) The regression plot of the SVR model for the A357-T6 cast aluminum alloy. (b) Kitagawa Diagram of the cast aluminum alloy A357-T6 under-tension loading, R= -1, using the SVR model.

The SVR model, applied to A357-T6 and trained on A356-T6 data, accurately predicts the Kitagawa diagram under fully reversed tension (R = −1), as illustrated in Fig.6b. The model's ability to be applied to related cast aluminum alloys is demonstrated by a moderate correlation (R^2 =0.738), successfully capturing the defect–fatigue strength relationship with an RMSE of 3.35 MPa.

RESULTS

Two statistical criteria were used to evaluate the model's prediction ability: R^2 (squared) and the RMSE.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\left(z_i - \hat{z}_i\right)^2} \quad , \quad R^2 = 1 - \frac{\sum\left(z_i - \hat{z}_i\right)^2}{\left(z_i - \bar{z}\right)^2}$$

$\hat{z}_i$ is the prediction, z_i is the true value, n is the number of samples, $\bar{z}$ is the average of z_i.

Tab. 5 is a summary of the multiple comparison test findings.

According to the summary of multiple comparison test findings in Tab. 5, the SVR model demonstrated the highest predictive accuracy for the A356-T6 alloy. It achieved a coefficient of determination R^2 of 0.957 and the lowest RMSE of

1.97 MPa. This indicates that the SVR model can accurately capture the fatigue strength transition between defect-dominated and defect-insensitive regimes.

Alloy type	ML model	R^2	RMSE [MPa]
A 356-T6	SVR	0.957	1.97
	RF	0.956	1.98
	GPR	0.858	3.58
A 357-T6	SVR	0.738	3.35

Table 5: Prediction accuracy of different machine learning methods for the two cast aluminum alloys.

While other models were effective, they showed slightly lower performance metrics:

• The RF model showed good accuracy with an R^2 of 0.956, but it had a slightly higher prediction error (RMSE of 1.98 MPa) compared to the SVR.

• Among the three models tested for A356-T6 in Tab. 5, the GPR model had the lowest performance, with an R^2 of 0.858 and a significantly higher RMSE of 3.58 MPa.

Tab. 5 also evaluates the global ability of the SVR model by testing it on the A357-T6 alloy (using a fixed SDAS value of 38 μm) after being trained on A356-T6 data. The model achieved a "moderate" predictive ability with an R^2 of 0.738 and an RMSE of 3.35 MPa. This demonstrates the model's robustness and its ability to successfully capture the defect–fatigue limit relationship across related cast aluminum alloys.

The high accuracy shown in Tab. 5 is directly attributed to the hybrid empirical–machine learning framework. Because experimental data was scarce, an empirical equation is used to generate a synthetic dataset of 5000 points. This augmented training set allowed the machine learning models to learn the fundamental physical correlations between SDAS, defect size, and fatigue limits more effectively than using limited experimental points alone.

Discussion

It is evident from the comparative evaluation of the investigated models for the A356-T6 alloy that the SVR approach is superior. Tab. 5 shows that this model had the highest predictive accuracy with a coefficient of determination R^2 of 0.957 and a low RMSE of 1.97 MPa.

The model's ability to precisely depict the nonlinear fatigue strength transition within the Kitagawa diagram is a major scientific contribution of this work. Without overfitting data fluctuations, the SVR and RF models effectively capture the physical transition from the microstructure-dominated (defect-insensitive) regime to the defect-dominated regime. This framework offers a reliable tool for microstructure-informed design optimization by acting as a physics-informed surrogate model, demonstrating that ML can surrogate and generalize universal metallurgical interactions across various cast aluminum systems.

The ML framework incorporates two primary physical drivers to address the underlying metallurgical mechanisms:

• As a microstructural parameter governed by cooling rates during solidification, SDAS serves as the primary determinant of fatigue limit in defect-free regions. The ML model uses SDAS to represent the intrinsic resistance of the aluminum matrix.

•By using the Murakami parameter, the model takes into account the local stress concentration generated by gas pores or shrinkage cavities. These defects act as crack initiation sites.

The success of this predictive framework is directly attributed to the hybrid empirical–machine learning strategy. By generating a large synthetic dataset of 5000 points based on a calibrated empirical equation, the study ensures that the fundamental physical correlations between SDAS, defect size, and fatigue limit are strictly maintained during the training process. In this context, the ML algorithm does not function as a "black box" but rather as a surrogate model designed to reproduce the empirical fatigue relationship under controlled conditions.

Furthermore, the model's robust aspect and global ability were validated by training SVR model on the A356-T6 alloy and testing it on the cast aluminum alloy A357-T6. Despite the variations in chemical composition, the model achieved an R^2 of 0.738 and an RMSE of 3.35 MPa for a fixed SDAS value of 38 μm. This demonstrates that the framework can successfully extrapolate the learned metallurgical interactions, specifically the competition between matrix strength (SDAS) and defect severity $\sqrt{area}$ to comparable cast aluminum systems.

In summary, the integration of microstructure-informed features allows the ML models to provide a scientifically grounded assessment of fatigue behavior, moving beyond simple data fitting to offer a reliable tool for microstructure-informed design optimization.

CONCLUSIONS

To predict the Kitagawa diagram of the cast aluminum alloy A356-T6, machine learning models, such as SVR, RF, and GPR, are created and trained for a range of SDAS values. Since experimental data is often scarce, an empirical equation is used to build a large synthetic dataset. With an RMSE of 1.97 MPa and $R^2 = 0.957$ for the A356-T6 alloy, the SVR model demonstrated the best performance. Second, a machine learning model is tested on the alloy A357-T6 after being trained with a fixed SDAS value of 38 μm on the cast aluminum A356-T6. The SVR model achieved an $R^2 = 0.738$, which reflects a good predictive accuracy.

Despite the high predictive accuracy, this study has limitations. First, the framework relies on a synthetic dataset generated via an empirical equation to overcome experimental data scarcity. Thus, the ML models function as surrogate models reflecting the underlying empirical logic. Second, the SDAS range is restricted to [25–80] μm, corresponding to standard casting conditions but potentially limiting extrapolation. Finally, the model focuses on defect size ($\sqrt{area}$) for spherical pores, without accounting for the complex morphology or spatial distribution of defects, which are critical for a comprehensive fatigue assessment.

Future research could investigate a number of approaches to enhance fatigue limit prediction. The creation of hybrid models, which integrate machine learning algorithms with experimental or physics-based methods to improve prediction accuracy, is one crucial path. Expanding the database to include a wider range of loading conditions and defect sizes would further enhance the model's generalization ability. Furthermore, the intricate nonlinear interactions between microstructural factors and the fatigue limit may be captured using deep learning techniques like deep neural networks (DNNs).

REFERENCES

[1] Tebaldini, M., Petrogalli, C., Donzella, G., Gelfi, M., La Vecchia, G. M. (2018). A356-T6 wheels: Influence of casting defects on fatigue design. Fatigue & Fracture of Engineering Materials & Structures, 41(1), pp. 1784–1793. DOI: https://doi.org/10.1111/ffe.12820.

[2] Shabani, M. O., Mazahery, A., Bahmani, A., Davami, P., Varahram, N. (2011). Solidification of A356 Al alloy: Experimental study and modeling. Kovove Materialy, 49, pp. 253–258. DOI: https://doi.org/10.4149/km_2011_4_253.

[3] Gawert, C., Bähr, R. (2021). Automatic Determination of Secondary Dendrite Arm Spacing in AlSi-Cast Microstructures. Materials, 14(11), 2827. DOI: https://doi.org/10.3390/ma14112827.

[4] Iben Houria, M. I., Nadot, Y., Fathallah, R., Roy, M., Maijer, D. M. (2015). Influence of casting defect and SDAS on the multiaxial fatigue behaviour of A356-T6 alloy including mean stress effect. International Journal of Fatigue, 80, pp. 90–102. DOI: https://doi.org/10.1016/j.ijfatigue.2015.05.012.

[5] Sun, J., Le, Q., Fu, L., Bai, J., Tretter, J., Herbold, K. (2019). Gas Entrainment Behavior of Aluminum Alloy Engine Crankcases during the Low-Pressure-Die-Casting Process. Journal of Materials Processing Technology, 266, pp. 274–282. DOI: https://doi.org/10.1016/j.jmatprotec.2018.11.016.

[6] Wang, R., Wu, S., Chen, W. (2018). Mechanism of burst feeding in ZL205A casting under mechanical vibration and low pressure. Transactions of Nonferrous Metals Society of China, 28, pp. 1514−1520. DOI: https://doi.org/10.1016/S1003-6326(18)64792-2.

[7] Murakami, Y. (2002). Metal Fatigue: Effects of Small Defects and Non-metallic Inclusions. Elsevier, Oxford. DOI: https://doi.org/10.1016/B978-0-08-043586-4.50023-5.

[8] Karash, E. T. (2022). The Effect of Stress Ratio on Fatigue Cracks Growth Rate in Aluminum Alloy. WSEAS Transactions on Applied and Theoretical Mechanics, 17, pp. 235–244. DOI: https://doi.org/10.37394/232011.2022.17.28.
[9] Thomas, A., Durmaz, A. R., Alam, M., Gumbsch, P., Sack, H., Eberl, C. (2023). Materials fatigue prediction using graph neural networks on microstructure representations. Scientific Reports, 13, 12562. DOI: https://doi.org/10.1038/s41598-023-38947-2.
[10] Chen, J., Liu, Y. (2021). Probabilistic physics-guided machine learning for fatigue data analysis. Expert Systems with Applications, 168, 114316. DOI: https://doi.org/10.1016/j.eswa.2020.114316.
[11] Lian, Z., Li, M. and Lu, W. (2022). Fatigue life prediction of aluminum alloy via knowledge-based machine learning. International Journal of Fatigue, 157, 106716. DOI: https://doi.org/10.1016/j.ijfatigue.2021.106716.
[12] Utpat, V. S. and Kulkarni, S. G. (2023). Analysis of various machine learning algorithms for cast aluminium alloy to estimate fatigue strength. Journal of The Institution of Engineers (India): Series D, 104(1), pp. 61-70. DOI: https://doi.org/10.1007/s40033-022-00381-7.
[13] Che, Z., Peng, C. (2024). Improving Support Vector Regression for Predicting Mechanical Properties in Low-Alloy Steel and Comparative Analysis.Mathematics, 12, 1153. DOI: https://doi.org/10.3390/math12081153.
[14] Bhat, S. K., Ranjan, A., Upadhyaya, Y. S., Managuli, V. (2025). Fatigue strength prediction of Cobalt alloys using material composition and monotonic properties: ML-based approach.Materials Research Express, 12, 046505. DOI: https://doi.org/10.1088/2053-1591/adc5c8.
[15] Esmaeili, H., Avateffazeli, M., Haghshenas, M., Rizvi, R. (2025). A Hybrid Framework for Characterizing and Benchmarking Fatigue S-N Curves in Aluminum Alloys by Integrating Empirical and Data-Driven Approaches Fatigue & Fracture of Engineering Materials & Structures, 48, pp. 44–59. DOI: https://doi.org/10.1111/ffe.14459.
[16] Chen, S., Bai, Y., Zhou, X. (2024). A deep learning dataset for metal multiaxial fatigue life prediction. Sci Data 11, 1027. DOI: https://doi.org/10.1038/s41597-024-03862-4.
[17] Hills, M., Becker, T. H. (2025). Machine Learning-Based Prediction of Fatigue Strength in Additively Manufactured Ti–6Al–4V Parts: A Sensitivity Analysis of Input Features, Journal of Materials Science & Technology.
[18] Shin, H., Yoon, T., Yoon, S. (2025). Fatigue life predictor: predicting fatigue life of metallic material using LSTM with a contextual attention model, RSC Advances, 15, 15781 15795. DOI : https://doi.org/10.1039/D5RA01578B.
[19] Zhang, Q., Dai, Y., Yue, H., Guo, C., Ji, Z., Li, Q., Zhang, J. (2025). Materials, 18(2), 332. Fatigue Life Prediction of 2024 T3 Clad Al Alloy Based on an Improved SWT Equation and Machine Learning, Materials, 18(2), 332. DOI: https://doi.org/10.3390/ma18020332.
[20] Li, C., Kapure, T. S., Roy, P. C., Gan, Z., Shen, B. (2025). DeepOFormer: Deep Operator Learning with Domain informed Features for Fatigue Life Prediction, arXiv. arXiv :2503.22475 pp. 1–6.
[21] Murakami, Y. (1993). Metal Fatigue : Effects of Small Defects and Nonmetallic Inclusions. Elsevier, Oxford, pp. 13–23. DOI: https://doi.org/10.1016/B978-0-08-042841-5.50006-8.
[22] Mu, P., Nadot, Y., Nadot-Martin, C., Chabod, A., Serrano-Muñoz, I., Verdu, C. (2014). Influence of casting defects on the fatigue behavior of cast aluminum AS7G06-T6. International Journal of Fatigue, 63, pp. 97-109. DOI: https://doi.org/10.1016/j.ijfatigue.2014.01.011.
[23] Wang, Q. G., Apelian, D., Lados, D. A. (2001). Fatigue behavior of A356/357 aluminum cast alloys. Part II- effect of microstructural constituents. Journal of Light Metals, 1, pp. 85–97. DOI: 10.1016/S1471-5317(00)00009-2
[24] Ben Ahmed, A., Nasr, A., Fathallah, R. (2017). International Journal of Advanced Manufacturing Technology, 90, pp. 3275–3288. DOI: https://doi.org/10.1007/s00170-016-9535-8.
[25] Vapnik, V., Golowich, S. E., Smola, A. (1996). Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing· Advances in Neural Information Processing Systems, 9, pp. 281- 287.
[26] Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, pp. 199–222. DOI: https://doi.org/10.1023/B:STCO.0000035301.49549.88.
[27] Breiman, L. (2001). Random Forests. Machine Learning, 45, pp. 5–32. DOI: https://doi.org/10.1023/A:1010933404324.
[28] Williams, C. K. I. and Rasmussen, C. E. (1995). Gaussian processes for regression. In Advances in Neural Information Processing Systems 8, pp. 514-520. MIT Press.