5-2-2023

Utilization of Machine Learning to Investigate Material State

Ana B. Abadie
Georgia Southern University

Utilization of Machine Learning to Investigate Material State

An Honors Thesis submitted in partial fulfillment of the requirements for Honors in *Mechanical Engineering.*

By
Ana Abadie

Under the mentorship of *Dr. Hossain Ahmed*

ABSTRACT

The ability to predict material behavior that undergoes various loading conditions is critical to the development of reliable and safe components. Thermal and mechanical fatigue loading can cause significant damage to materials, leading to failure and potential safety hazards. Machine learning algorithms have emerged as a promising tool for improving accuracy and efficiency of predicting material behavior under such loading conditions. This research provides a comprehensive overview of a machine learning algorithm that is able to analyze and predict material state independently of the loading sequence. Unidirectional carbon fiber reinforced polymer (UD CFRP) composite which has undergone two different loading sequences is analyzed. On one hand, the pristine material sustains 40 cycles of thermal fatigue followed by a break where data is acquired and then the specimen undergoes 150k cycles of mechanical fatigue. The second load sequence consists of the same cycles but in a reversed manner. Data collected from a small section of the composite in each stage was used to train a Naive Bayes classifier. A feature selection process is carried out with the use of principal component analysis to identify the most relevant parameters for use.The results show that the Naive Bayes classifier can accurately predict the fatigue sequence of UD CFRP samples under thermal and mechanical loading conditions with an accuracy of up to 80%. The findings of this paper can have significant implications for the analysis of structures subjected to thermal and fatigue loading.

Thesis Mentor: *Dr. Hossain Ahmed*

Honors Dean: Dr. Steven Engel

May 2023
Department of Mechanical Engineering
Honors College
Georgia Southern University

Acknowledgements

I would like to express my deepest gratitude to my thesis mentor Dr. Hossain Ahmed for his support, guidance, and encouragement throughout my research journey. Your insightful comments, constructive criticism and patience have greatly contributed to the success of this project. I'm also grateful to Georgia Southern University's Mechanical Engineering Department for providing the necessary courses and resources for my research.

I want to thank my parents Francisco Abadie, Ana Lucia Molina and my grandmother Toty Abadie for their unwavering support, love and encouragement throughout my college career. Without them, this would not have been possible. You have all been my greatest motivation even from a long distance.

Thank you all for the support and guidance which has made this journey an unforgettable experience.

Table of Contents

1. Introduction

1.1 Background and Motivation

Unidirectional Carbon Fiber Reinforced Polymer (UD CFRP) is the material of study for this research as shown in Figure 1. This composite is increasingly used in a wide range of industries due to their high strength-to-weight ratio, corrosion resistance and other desirable mechanical properties. Industries that commonly use this material include aerospace and aviation, automotive, wind energy, civil engineering, medical and many more. One of the key advantages of CFRP composites is their high temperature resistance. They have a relatively low coefficient of thermal expansion ranging from -1K^{-1} to 8K^{-1}[1], therefore exhibit minimal dimensional changes when exposed to temperature fluctuations. This property helps prevent issues such as warping, distortion or cracking. In aerospace applications, where an aircraft experiences significant temperature variations during operations as well as changes in pressure, UD CFRP composites are preferred due to their ability to maintain their structural integrity. Additionally, they have excellent resistance to environmental factors such as moisture, chemicals, and UV radiation, making them highly durable. For example, the most common applications of UD CFRP composites in aircraft structures include components such as wings, fuselages, and tail sections [2]. In addition to aircraft structures, they can be used in space applications such as satellites and space probes and more specifically the solar panels on the Mars Reconnaissance Orbiter [3].

Figure 1. Unidirectional Carbon Fiber Reinforced Polymer

Understanding material states is important in engineering and materials science because it allows for prediction and control of material behavior under different conditions. Material states refer to the physical and mechanical properties of a material, such as its composition, microstructure, and external loading conditions, which influence its behavior. It is essential for designing and optimizing material properties for specific applications. Understanding material state is crucial for predicting and analyzing material failure and material characterization. This in turn helps engineers optimize manufacturing processes for a particular material. By controlling processing conditions such as temperature, pressure and time, engineers can tailor materials to achieve desired properties for any given application. However, estimation of material states under various types of externally applied loads and corresponding change of material properties are challenging to determine. In this respect analytical, numerical, and experimental research has been conducted by researchers and scientists from academia and their industry partners [4]. Recently, utilization of machine learning algorithms on experimental data is becoming one of the major tools in engineering [5]. In this study, a customized machine learning algorithm has been proposed to estimate the material state of transversely isotropic unidirectional composite structures.

Machine learning has been widely leveraged in discovering new functional material. It has become increasingly popular in materials science research due to their ability to extract complex relationships between material properties and processing conditions. Traditionally, material scientists have relied on experimental techniques to analyze the behavior of materials, such as microscopy, spectroscopy, and mechanical testing [6]. These methods can be time-consuming and costly, and they may not provide a complete understanding of the material's behavior. By analyzing large datasets of material properties, machine learning algorithms can identify new materials with desired characteristics. Furthermore, machine learning models can predict material properties based on their chemical and physical characteristics, eliminating the need for costly experimental testing [7]. Machine learning algorithms can also enhance quality control by analyzing data from manufacturing processes to detect defects or anomalies in materials [8]. Machine learning can analyze data from failed materials to identify their root cause of malfunctioning, giving engineers insight into how to enhance design and manufacturing processes [9]. Finally, machine learning can optimize manufacturing processes by analyzing data and recognizing opportunities to increase efficiency and reduce waste [10].

Overall, machine learning holds the potential to revolutionize how material state investigation is done in engineering industries. Benefits of using machine learning to investigate material state include improved efficiency to process large amounts of data in a relatively short amount of time [11], enhanced predictive capabilities based on various input parameters, reduced experimental costs, and enhanced process optimization. Moreover, the demand for advances and smart engineering materials has increased. In this research, finding the correct tool to analyze a large dataset was explored to understand material state. Data collected from a small section of the composite in each stage can be used to train an algorithm to validate material recovery and memory. In this way, Machine Learning techniques are used to analyze and predict material state independently of the loading sequence applied.

1.2 Principal Component Analysis

Principal component analysis (PCA) is a multivariate technique that analyzes data in which observations are described by several inter-correlated quantitative variables [12]. In the context of material state investigation, principal component analysis was applied to extract relevant features from the scanning acoustic microscopy (SAM) data. SAM scans were taken at six different stages of each specimen's life to monitor its internal conditions. These SAM scans result in high-dimensional datasets, making it challenging to extract meaningful information from them directly. PCA can be used in similar applications to reduce the dimensionality of these datasets by identifying the most important variables that explain the most significant variance in the data.

With the implementation of PCA, it is possible to identify principal components and identify patterns in the SAM data. Each principal component is a linear combination of the original variables that represent a particular pattern in the data. The first PC explains the most significant amount of variance, followed by the second PC, and so on. These components are ranked in order of importance and the ones that contribute the most to the data's variance can be used as features for further analysis. By reducing the dimensionality of the data, PCA helps overcome the issues like overfitting and poor generalization performance which arise from dealing with high-dimensional datasets. Moreover, it becomes possible to identify and remove highly correlated variables, which can help overcome the problem of multicollinearity [13]. By combining correlated variables into principal components orthogonal to one another, PCA helps improve accuracy and stability of regression models.

1.3 Naive Bayes

In materials science, machine learning techniques have been employed for tasks such as materials discovery, property prediction, and structure-property relationships. One notable method is the Naive Bayes classifier - an efficient probabilistic algorithm used widely in text classification and spam filtering applications [14]. They rely on Bayes' theorem, which states that the probability of a hypothesis (or class) given evidence (or input features) is proportional to its prior probability multiplied by that evidence's probability. Naive Bayes classifiers work by calculating the probability of each class given input features, then selecting the one with highest probability as predicted class. To do this, they estimate prior probabilities for each class and conditional probabilities for each feature given each class using a training dataset of labeled examples.

The "naive" part of this algorithm's name comes from its assumption that all input features are independent from one another, making conditional probability calculations easier [15]. Although this assumption may not always be correct, the algorithm still performs well in practice. Naive Bayes classifiers are commonly employed for text classification tasks such as spam filtering and sentiment analysis, but they can also be employed for image classification, medical diagnosis, and credit risk assessment. Naive Bayes classifiers have the advantage of high-dimensional datasets with many input features where other algorithms may require too much computational power or risk overfitting.

Materials science and engineering rely on the study of materials to understand their properties. Material behavior under various environmental conditions and external stimuli plays a major role in determining its suitability for various applications. Traditionally, experimental methods have been employed to investigate material state and properties;

however, recent advances in machine learning techniques have opened new opportunities to utilize computational methods for material state investigation.

This thesis investigates the application of a Naive Bayes classifier for material state analysis. The goal is to create an algorithm that can accurately predict a material's state based on its signal properties, trained on a dataset with known states, then tested against another dataset to confirm its accuracy and dependability.

While other approaches may be suitable for finding material states, Naive Bayesian methods and principal component analysis are often used in materials science research due to their simplicity, interpretability, and ability to handle large datasets. Some machine learning models can be complex and difficult to interpret as well as computationally expensive, making it difficult to understand the relationship between material properties and processing conditions. On the other hand, certain algorithms require the selection of relevant features which can be time consuming and challenging. Large amounts of data are necessary for most machine learning models in order to provide accurate predictions. In contrast, Naive Bayes and PCA are both computationally efficient and provide meaningful results that can be easily understood, even with limited data. Ultimately, the selection of an appropriate approach depends on the specific research question and data availability.

This research has the potential to make significant advances in materials science and engineering. By creating an accurate model for investigating material state, we can expedite discovery of new materials and optimize their properties for various applications. Furthermore, using machine learning techniques for material state investigation can result in significant time and cost savings when compared to traditional experimental methods.

Overall, this thesis presents a novel approach for investigating material state using machine learning techniques, particularly the Naive Bayes classifier and principal component analysis. The combination of Naive Bayes and PCA has been shown to be effective in analyzing material data. For example, the combination of these techniques was used to predict the formation energy of metal-organic frameworks [16] and classify alloys based on their microstructure [17]. By developing and evaluating this model, we hope to demonstrate the potential of machine learning in materials science and engineering and pave the way for further research in this rapidly expanding area.

2. Experimental Methodology

2.1 Development of Experimental Data

A set of six unidirectional (UD) carbon fiber reinforced polymer (CFRP) composite specimens were created and polished to expose the UD fibers. The specimens were prepared according to ASTM D3039 standard. The specimens were then divided into two groups, with each group consisting of three specimens. Group 1 underwent mechanical fatigue loading using an MTS tensile tester machine and then thermal loading using an oven and ice-bucket system. On the other hand, Group 2 underwent thermal fatigue loading followed by mechanical loading using the same procedures as Group 1 as shown in Figure 2.

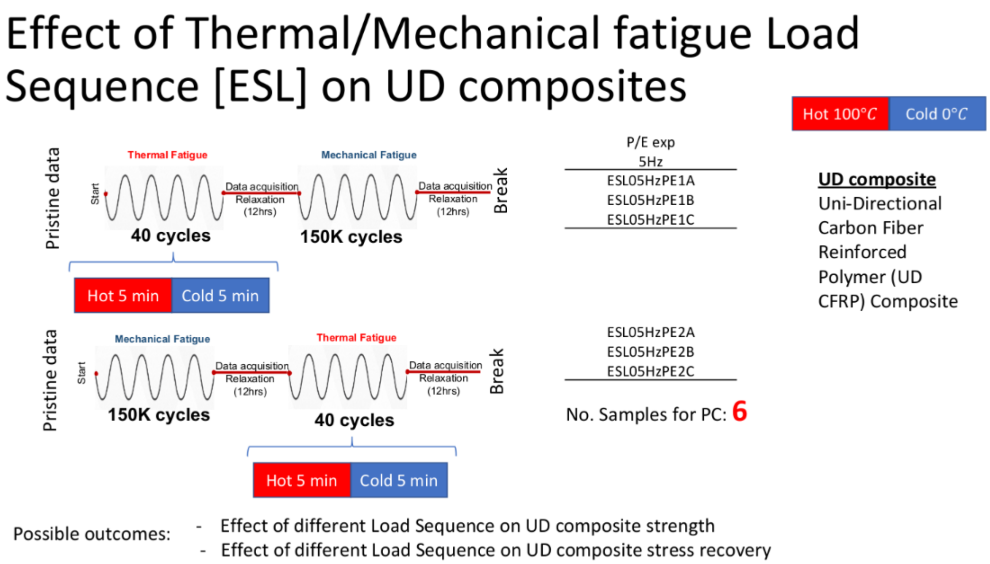

Figure 2. Effect of Thermal/Mechanical Fatigue Load Sequence on UD Composites

The mechanical loading was accomplished by cyclic tensile fatigue at a frequency of 5 Hz for a total of 40,000 cycles at 60% of the material's ultimate tensile strength, which was found to be 11,500N. Meanwhile, the thermal loading was performed by subjecting the specimens to 100℃ for three minutes, then immersing them in ice water for another 3 minutes, for a total of 150 cycles.

Scanning Acoustic Microscopy (SAM) scans were taken of the specimens at three points during each specimen's life which include the pristine state, the 12-hour relaxation period after each loading cycle. Therefore, the first group of specimens went through SAM scan after thermal fatigue cycles and mechanical fatigue cycles. On the other hand, the second group of specimens went through SAM scans after thermal fatigue cycles and then after mechanical fatigue cycles. A 12-hour window was imposed between the thermal and mechanical cycles of each specimen to ensure the samples were adequately rested and alert.

The reason behind altering the loading cycles was to intentionally expose the specimens to two different types of material states.

2.2 DATA ANALYSIS USING MATLAB

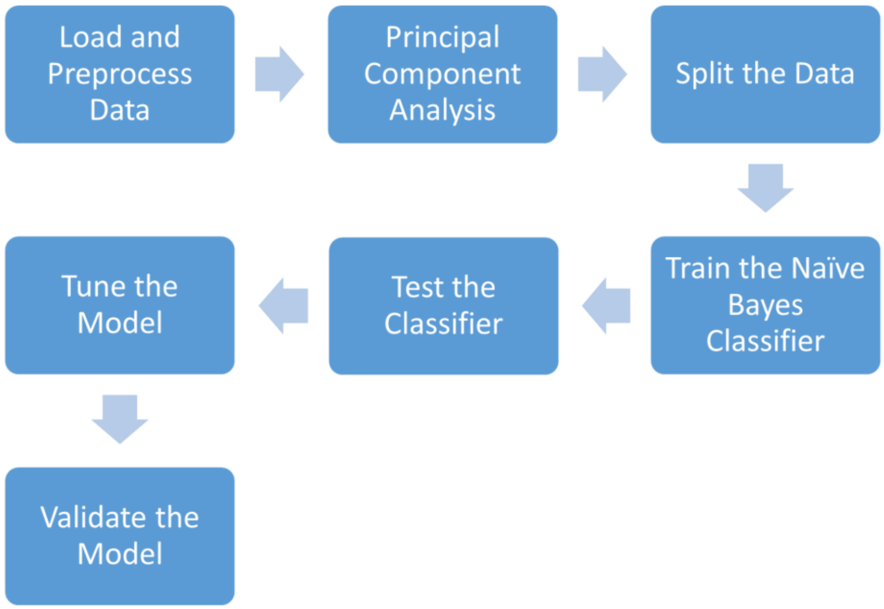

Figure 3. Data Analysis Flowchart

With the use of MATLAB, code was generated to perform the operations as shown in Figure 3. First, data from multiple .SAZ files which contain ultrasonic signals are read and important features from these signals are extracted at specific points in the region of interest. More specifically, it reads all .SAZ files from a selected folder and performs calculations to extract features from each signal. The code sorts the .SAZ files in chronological order and determines the total number of files. It then initializes several variables that store the extracted features.

For each .SAZ file, the code reads in the signal point by point and applies various processing techniques to extract features such as nonlinearity, Hilbert envelope, maximum frequency and amplitude, and instantaneous energy. The extracted features are then stored in a matrix called ‘data’, which also includes a column indicating the index of the .SAZ file to which the feature belongs.

Once all the signals have been processed for a given .SAZ file, MATLAB calculates descriptive statistics on the extracted features including minimum, maximum, median, and mean values. These statistics, along with the index of the .SAZ file, are stored in another matrix called ‘FData’. This matrix is saved in a .mat file with a name based on the index of the signal file. This is repeated for all files in a selected folder, resulting in a set of .mat files each containing feature data for a single .SAZ file.

Then, a separate MATLAB code performs data analysis and classification tasks using PCA and Naive Bayes classifier. First, it loads data from Group 1 (TM), Group 2 (MT) and Pristine state files and stores them in variables using the ‘sprintf’ function to dynamically generate the file names based on the loop index ‘i’. Next, the code concatenates the data from Group 1, Group 2 and Pristine into a single matrix ‘X’ and creates a corresponding vector ‘Y’ to label each data point as belonging to either Group 1, Group 2 or Pristine.

Afterwards, principal component analysis is performed on the data in ‘X’ using the ‘pca’ function. The function calculates the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors are the principal components of the data and the eigenvalues represent the amount of variance captured by each PC. It then calculates the cumulative explained variance of each principal component and determines the number of principal components to use in the analysis by selecting the smallest number of components needed to explain at least 95% of the variance. It then projects the data onto the selected principal components.

The code then splits the data into training and testing sets and trains a Naive Bayes classifier using the ‘fitcnb’ function with the training data. The ‘fitcnb’ function trains a Gaussian Naive Bayes classifier, where each class is modeled by a Gaussian distribution with mean and variance estimated from the training data. It tests the classifier using the testing data and computes performance measures such as accuracy, precision, recall and F1 score using the ‘confusionmat’ function and a for loop. Finally, the results of the classification in the variables are stored in the variable ‘confusion_matrix’. The confusion matrix shows the number of true positives, false positives, true negatives, and false negatives for each class. These performance measures are then used to tune and validate the model.

3. Results

After conducting extensive analysis on the large set of signal data, it was found that the Naive Bayes classifier trained on PCA processed data was able to predict the loading sequence of the samples with an accuracy of up to 80%. This was determined using a confusion matrix which indicated that out of the 21 observations in the testing set, 17 were predicted correctly and 4 were predicted incorrectly as shown in Figure 4. This indicates that the algorithm struggled to distinguish between groups 1 and 2 (MT & TM), with three observations being misclassified between these two groups. Additionally, one observation in the pristine group was misclassified.

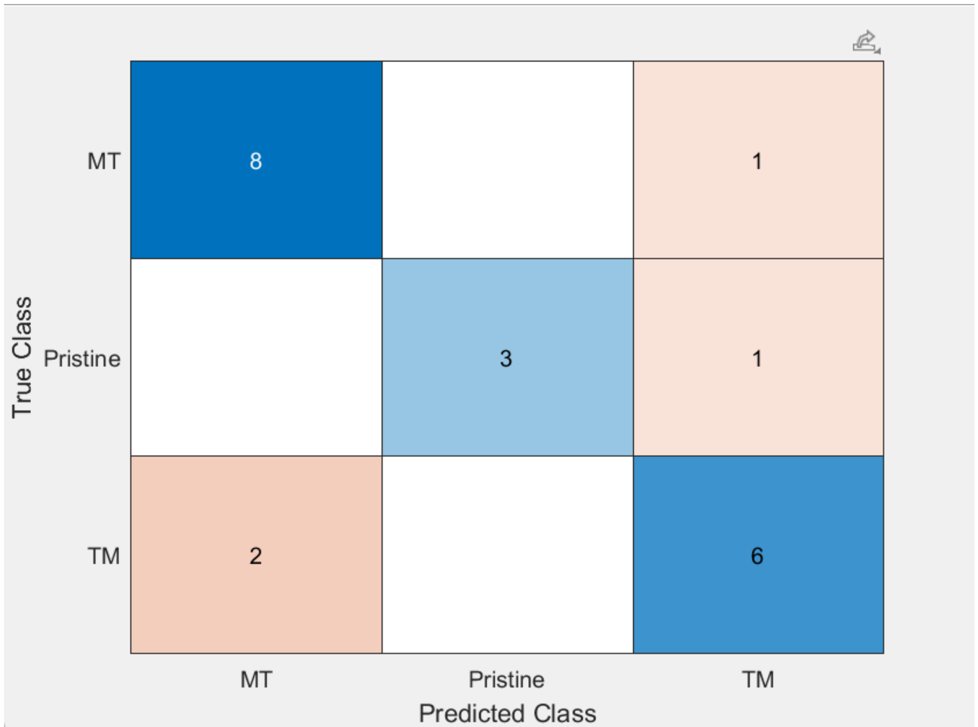

Figure 4. Confusion Matrix

The precision and recall scores for each group provide further insight into the algorithm's performance. Precision is the proportion of true positive predictions out of all the actual positive instances. Recall is the proportion of true positives over the total number of actual positives. For Group 1, the precision score was 0.8, indicating that 80% of the instances predicted in that group actually belonged there. The recall score for Group 1 was 0.89, indicating that the algorithm correctly identified 89% of the instances in Group 1. For Group 2, the precision score was 0.86 and recall score was 0.75. For the pristine group, the recall score was 0.75 and precision score was 0.75.

4. Discussion

The algorithm was able to predict with an overall accuracy of 80%, which is a relatively good performance. However, the misclassifications of three observations between group 1 and 2, and on with pristine, indicates that there might be some overlap or similarity between these groups. This can be further explored with additional data or by modifying the features used in analysis.

The precision and recall scores for each group show that the algorithm performed well for all groups, given that it was above 0.75 for all of them. A precision score of 1 would indicate that there are no false positives, while a recall score of 1 means that there were no false negatives. In this case, a precision score of 1 was not achieved for any group, indicating there were misclassifications.

In many cases, it is difficult if not impossible to distinguish between different loading sequences that a material undergoes. When it comes to validating material recovery and memory, it is important to ensure that the material is able to recover its original properties after undergoing thermal or mechanical fatigue loading. UD CFRP composites were unloaded and allowed to rest for a certain amount of time, after which it is reloaded, and signal data is measured again. If the material is able to recover its original properties after unloading and reloading, it can be said to exhibit material memory.

Overall, the successful classification of loading sequences using machine learning algorithms can help in the validation of material recovery and memory. The results of this analysis demonstrate the potential use of machine learning algorithms, such as Naive Bayes, to classify and distinguish between different loading sequences. Further research can explore the use of other algorithms or feature selection techniques to improve the accuracy of the classification. It can also lead to the development of more advanced materials with tailored mechanical properties that can withstand specific loading conditions.

5. Conclusion

The present study aimed to investigate machine learning algorithms, specifically PCA and Naive Bayes, to distinguish between different loading sequences applied to Unidirectional Carbon Fiber Reinforced Polymers. The results indicate that the algorithm was able to predict with an overall accuracy of 80%, with most misclassifications occurring between groups 1 and 2. Precision and recall scores were calculated for each group. Scores of 0.75 or above were obtained for both recall and precision, providing a more comprehensive evaluation of the algorithm's performance.

Overall, the findings of this study suggest that machine learning algorithms have the potential to become a valuable tool for analyzing large datasets in materials science applications. By accurately distinguishing between different loading sequences, the algorithm can be used to improve the efficiency and effectiveness of composites undergoing recovery processes, leading to cost savings and reduced environmental impact.

There are several limitations to the present study that should be addressed in future research. For instance, the dataset used in this study was relatively small and more data may be needed to increase accuracy and generalizability of the algorithm. Additionally, the algorithm was only implemented on two specific loading processes and its applicability to other processes remains unclear. Future research could also explore the use of other machine learning algorithms or combinations of such to improve the efficiency and accuracy of the model. Fine-tuning of the hyperparameters of the Naive Bayes classifier and PCA method can also optimize its performance. While further research is needed to fully realize this potential, the findings of this study highlight the importance of continuing to explore the use of machine learning in material state investigation.

6. References

[1] Joven, R. (2012, January 12). *Study of Thermal Expansion in Carbon Fiber Reinforced Polymer Composites*. ResearchGate. https://www.researchgate.net/publication/262006977_Study_of_Thermal_Expansion_in_Carbon_Fiber_Reinforced_Polymer_Composites

[2] Aminabhavi, T. M., & Kulkarni, A. (2017). *Study of aerospace applications of advanced composites.* International Journal of Engineering and Techniques, 3(3), 59-63.

[3] Beumler, T., & Johnson, K. (2010). *Carbon Fiber Reinforced Polymer Composites in Aerospace Engineering.* Journal of Reinforced Plastics and Composites, 29(10), 1505-1522.

[4] Callister, W. D., & Rethwisch, D. G. (2018). *Materials science and engineering: An introduction*. Wiley.

[5] Rodrigues, J.F., Florea, L., de Oliveira, M.C.F. et al. (2021) *Big data and machine learning for materials science.* Discov Mater 1, 12 . https://doi.org/10.1007/s43939-021-00012-0

[6] Mangalampalli, S. R., & Liang, X. (2019). *Machine learning in materials science*. Journal of Materials Research, 34(22), 3741-3759. doi: 10.1557/jmr.2019.290

[7] Chibani, S., & Coudert, F.-X. (2020, August 4). *Machine learning approaches for the prediction of materials properties*. AIP Publishing. https://pubs.aip.org/aip/apm/article/8/8/080701/122765/Machine-learning-approaches-for-the-prediction-of

[8] Escobar, C. A., & Morales-Menendez, R. (2018, February). *Machine learning techniques for quality control in high conformance manufacturing environments*. ResearchGate. https://www.researchgate.net/publication/323448597_Machine_learning_techniques_for_quality_control_in_high_conformance_manufacturing_environment

[9] Biswas, S., Fernandez Castellanos, D. & Zaiser, M. (2020) *Prediction of creep failure time using machine learning*. Sci Rep 10, 16910 . https://doi.org/10.1038/s41598-020-72969-6

[10] Weichert, D., Link, P., Stoll, A. et al. (2019) *A review of machine learning for the optimization of production processes*. Int J Adv Manuf Technol 104, 1889–1902 . https://doi.org/10.1007/s00170-019-03988-5

[11] Smith, M.R., Martinez, E., and Johnson, K. (2020). *Machine learning algorithms for data analysis: A primer. Journal of Data Science* 18(2), 301-316.

[12] Abdi, H., & Williams, L. J. (2010, July 15). *Principal component analysis - abdi - 2010 - wiley online library*. Retrieved May 1, 2023, from https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.101

[13] Huang, J., Zhang, W., Feng, W., & He, H. (2019). *Feature selection and PCA-based extreme learning machine for fault diagnosis*. IEEE Access, 7, 20322-20329.

[14] Rish, I. (n.d.). *An empirical study of the naive bayes classifier - gatech.edu*. Retrieved from https://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf

[15] Mitchell, T. M. (2020, October 1). *Machine learning - Carnegie Mellon University*. Retrieved from https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf

[16] Kim, K., Lee, J. H., & Kim, S. (2017). *Predicting the formation energy of metal-organic frameworks by machine learning with effective feature descriptors and principal component analysis*. Journal of Chemical Information and Modeling, 57(7), 1694-1704.

[17] Chen, Z., Zuo, X., Zhang, X., & Chen, L. Q. (2018). *Microstructure classification of alloys using principal component analysis and naive Bayes*. Acta Materialia, 151, 222-229.