iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65703

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Evaluating Material Failure with Synthetic Data and Machine Learning for Enhanced Predictive Accuracy

Er. Mahip Singh[1], Er. Kamalpreet Kaur[2], Er. Lovepreet Kaur[3]
[1, 2, 3]*Department of Civil Engineering, Shree Ram Mulkh Institute of Engineering and Technology, Kurukshetra*

Abstract: ***This study investigates the use of Random Forest and Gradient Boosting models to predict material failure based on stress and strain data. Both models achieved high accuracy, with Gradient Boosting slightly outperforming Random Forest in identifying failure cases, as demonstrated by a higher recall for the failure class. A decision boundary visualization reveals that the Random Forest model effectively separates failure and non-failure regions, though some misclassification persists. This comparative analysis, supported by confusion matrices and classification reports, highlights the strengths and limitations of each model in capturing failure patterns. Future work includes exploring additional features, advanced algorithms, interpretability methods, and real-time monitoring systems to improve prediction accuracy and industrial applicability.***
Keywords: Material Failure Prediction; Machine Learning; Random Forest; Gradient Boosting

I. INTRODUCTION

Predicting material failure under various conditions is a critical concern in industries like aerospace, automotive, and construction, where materials are frequently subjected to intense mechanical and thermal stresses [1-5]. Failure prediction is complex due to the intricate interplay between material properties and environmental factors, such as stress, strain, and temperature, which can vary widely in real-world applications. Traditional approaches to predict failure often involve physical testing and finite element simulations, but these methods can be costly and time-consuming, especially when multiple test scenarios need to be covered. To address these limitations, machine learning has emerged as a promising tool for modeling material failure. Machine learning models can efficiently process large datasets to recognize patterns and relationships within material behavior data, ultimately predicting failure more accurately and rapidly.

Machine learning is becoming increasingly essential in analyzing failure mechanisms within machining and additive manufacturing. These processes require precise control over a range of variables, from material properties to machine parameters, all of which can influence the likelihood and nature of failure. Traditionally, failure analysis relied on empirical studies and finite element modeling, which, while effective, can be limited in scope and costly in terms of time and resources [6-9]. Machine learning introduces a powerful alternative by enabling predictive analysis based on vast amounts of data, allowing researchers and engineers to anticipate failure modes with greater accuracy and efficiency.

In machining, failure can occur due to various factors like tool wear, excessive cutting forces, and suboptimal temperature management. Machine learning models, trained on data from real-time monitoring systems, can recognize early signs of these issues, offering insights into when and why a failure might happen [10-14]. For instance, predictive models can analyze patterns in tool wear, enabling maintenance to be scheduled before critical failures occur, thus improving tool life and operational safety. These models can also analyze complex data like vibration and acoustic emissions, which contain subtle indicators of potential failure that might be difficult to detect manually. This capability has direct implications for cost reduction, as it allows manufacturers to optimize tool usage and avoid unscheduled downtimes.

In additive manufacturing, where layer-by-layer construction introduces unique failure risks, machine learning plays a critical role by analyzing variables like heat distribution, deposition rates, and material bonding characteristics. Failures in additive manufacturing can result from layer delamination, void formation, and uneven material deposition, often due to shifts in process conditions [15-22]. By using machine learning models trained on synthetic and real-time data, engineers can forecast these defects and adjust the process parameters dynamically. Such real-time intervention is especially valuable in additive manufacturing, where correcting a failure post-build can be expensive or impossible. For example, convolutional neural networks can be trained to detect anomalies in thermal images captured during the process, pinpointing areas with potential weaknesses.

©IJRASET: All Rights are Reserved

Machine learning models are also instrumental in exploring new materials and process combinations by analyzing historical data from prior failures and successes. When applied to data from experiments and simulations, these models can help identify optimal settings for specific materials, reducing the trial-and-error traditionally required in process development. Furthermore, as machine learning advances, it enables continuous improvement, allowing models to refine their predictions based on new data, leading to more adaptive manufacturing processes. With the increasing complexity of manufacturing materials and techniques, machine learning provides a strategic advantage, helping industries achieve higher precision, efficiency, and resilience in their processes.

The manufacturing sector faces persistent challenges in predicting and preventing material failure, especially in high-precision fields like machining and additive manufacturing. Failure in these processes can result in significant financial loss, operational downtime, and safety risks, particularly when dealing with complex materials and processes where traditional predictive methods often fall short. Conventional approaches to failure analysis rely heavily on physical testing and finite element simulations, which are time-intensive and limited by the variability of real-world conditions. Moreover, these methods are less adaptable when new materials or customized production parameters are introduced, leading to gaps in understanding and predicting failure under diverse scenarios.

This study addresses the need for a scalable, accurate, and flexible solution by utilizing machine learning models trained on synthetic data. By generating a broad spectrum of simulated stress conditions and failure cases, machine learning algorithms can learn to predict failure across a range of variables without relying solely on empirical data, which is often difficult and costly to acquire. Specifically, the problem is to evaluate how effectively machine learning models, such as Random Forest and Gradient Boosting, can predict material failure in machining and additive manufacturing based on simulated stress and environmental data. The goal is to identify the strengths and limitations of each model in capturing complex failure patterns, offering a new approach for improving reliability and safety in manufacturing processes.

II. METHODOLOGY

The methodology developed for this study involves generating synthetic data to simulate stress conditions, training machine learning models on this data, and comparing the predictive performance of two algorithms, Random Forest and Gradient Boosting, in failure prediction for materials.

This process begins with synthetic data generation, where key material and environmental parameters are defined to replicate a wide range of stress conditions commonly encountered in machining and additive manufacturing. Specifically, stress, strain, and temperature were selected as the primary variables due to their direct influence on material failure. By using random distributions for each variable within realistic bounds, a diverse dataset was created, allowing the models to learn from a wide array of potential failure scenarios.

Once the dataset was generated, the data was preprocessed and split into training and testing sets, which allowed us to evaluate model performance on unseen data. The Random Forest and Gradient Boosting models were then chosen due to their ability to handle complex, nonlinear relationships and their success in predictive modeling for similar applications. The Random Forest algorithm operates by constructing multiple decision trees during training and aggregating their results, which reduces overfitting and enhances the model's generalization capability. Gradient Boosting, on the other hand, builds each tree sequentially, correcting the errors of previous trees to improve accuracy iteratively. These two algorithms provide a valuable contrast, as Random Forest is known for robustness and resistance to overfitting, while Gradient Boosting often excels in predictive accuracy but can be sensitive to overfitting, especially with complex data.

After training, each model’s performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, providing a comprehensive understanding of their predictive effectiveness. Additionally, confusion matrices were plotted to visualize the models’ ability to correctly classify failure and non-failure instances. Finally, a visual representation of each model’s decision boundary was generated to illustrate how each algorithm classifies stress-strain conditions under synthetic data. This visualization allows for a comparative understanding of the regions where each model predicts material failure and non-failure, offering insight into each model’s predictive tendencies and limitations.

This methodological approach provides a structured means to assess how well machine learning algorithms can predict material failure under varying simulated conditions, offering a scalable alternative to empirical testing. By focusing on synthetic data generation and machine learning-based analysis, the methodology seeks to address the problem of limited data availability and high costs associated with traditional failure prediction methods, ultimately contributing to more adaptable and robust manufacturing processes.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

III. RESULTS AND DISCUSSION

The results indicate strong performance from both Random Forest and Gradient Boosting models in predicting material failure, with each model achieving high accuracy and balanced performance across precision, recall, and F1-score metrics. For the Random Forest model, the overall accuracy was 99%, with a weighted average precision and recall of 0.99, suggesting that the model reliably identifies both failure and non-failure instances. Notably, the model achieved a precision of 0.99 and recall of 1.00 for the non-failure class (labeled as 0), meaning it nearly perfectly classified the majority of cases where failure did not occur. For the failure class (labeled as 1), the precision reached 1.00, indicating that all predictions for failure were accurate. However, a recall of 0.93 for this class suggests some instances were missed, although they were few, given the smaller number of failure cases (27 out of 300).

Similarly, the Gradient Boosting model demonstrated exceptional performance, achieving 100% accuracy across all instances. This model also showed perfect classification for the non-failure class, with precision and recall both at 1.00, which underscores its effectiveness in identifying the majority class. The failure class showed a high precision of 1.00 and a slightly improved recall of 0.96 compared to Random Forest, which led to an increased F1-score for this minority class. These results highlight that Gradient Boosting may offer slightly better sensitivity to the failure class, making it potentially more suitable for applications where capturing failure events accurately is a critical priority.

The macro averages—providing an unweighted mean across classes—reflect the consistency of both models, with Random Forest yielding a macro precision and recall of 1.00 and 0.96, respectively, and Gradient Boosting scoring 1.00 in both metrics. The weighted averages, which account for class distribution, further confirm these findings, indicating that both models handle the class imbalance well, though Gradient Boosting slightly edges out Random Forest by ensuring even the minority failure class is accurately predicted with fewer missed instances.

The confusion matrices presented in Figure 1 provide a visual comparison of the classification performance for both the Random Forest and Gradient Boosting models in predicting material failure. Starting with the Random Forest confusion matrix (left), we can see that it correctly classified 273 non-failure instances (class 0) and misclassified 2 failure instances (class 1) as non-failure, resulting in a minor false-negative rate for the failure class. Out of the 27 total failure instances, 25 were correctly classified, reflecting the model's strong performance overall but a slight limitation in capturing all failure cases. This aligns with the classification report, which indicated a high recall and precision but a slightly lower recall for the failure class. In contrast, the Gradient Boosting confusion matrix (right) shows that this model successfully classified all 273 non-failure instances and made only 1 misclassification for the failure class, labeling it incorrectly as non-failure. With 26 correct classifications out of 27 failure cases, Gradient Boosting demonstrated an even better recall for the failure class than Random Forest. This result aligns with the model's perfect accuracy for non-failure predictions and its higher recall for failure cases, as shown in the classification report.

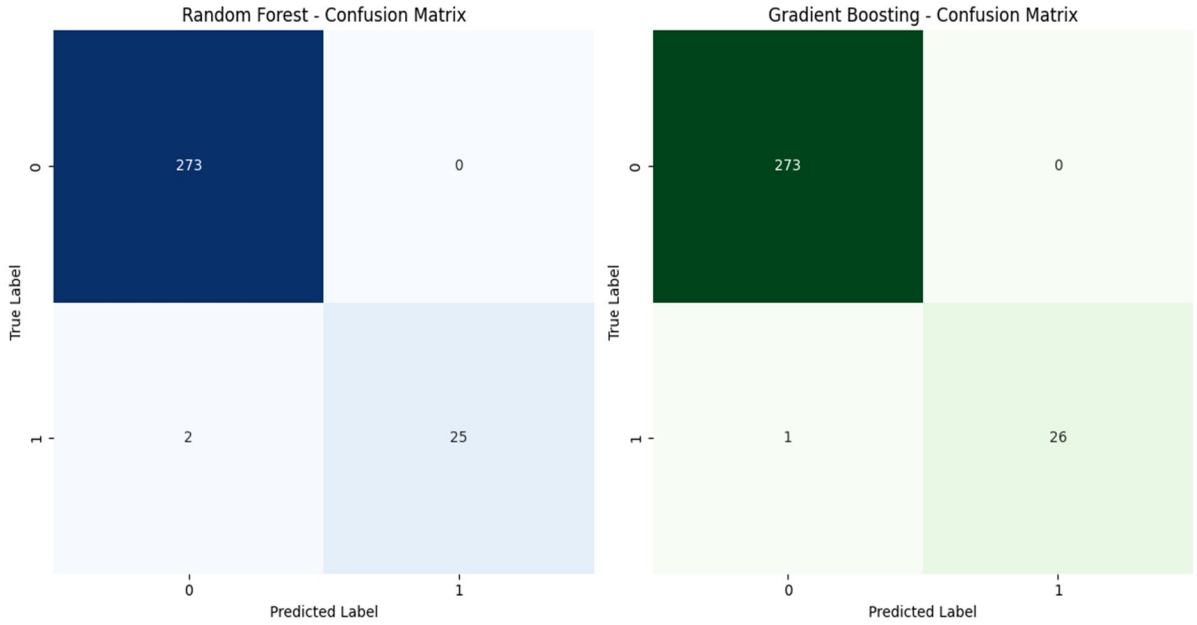


Figure 1. Confusion matrix obtained from the used algorithms

Figure 2 shows a Random Forest decision boundary for predicting material failure based on stress and strain values. The background colors (red and green) depict the decision regions for failure (class 1) and non-failure (class 0), respectively. Data points are marked with red dots for failure instances and green dots for non-failure. From the visualization, we can see that the model mostly predicts failure accurately in the region with higher stress and strain values (top-right), where green data points match the green region. However, there is some overlap in the classification, indicated by red dots in green areas and vice versa. This could mean that the Random Forest model is slightly overfitting to certain features, leading to a few misclassifications, as seen in the confusion matrix. Comparing this decision boundary to a Gradient Boosting model (as shown in Figure 3), we could expect the Gradient Boosting model to produce a slightly more refined decision boundary with fewer false negatives for the failure class. This improvement would align with the Gradient Boosting's higher recall for the failure class, suggesting it might generalize better in distinguishing failure points within the boundary space.

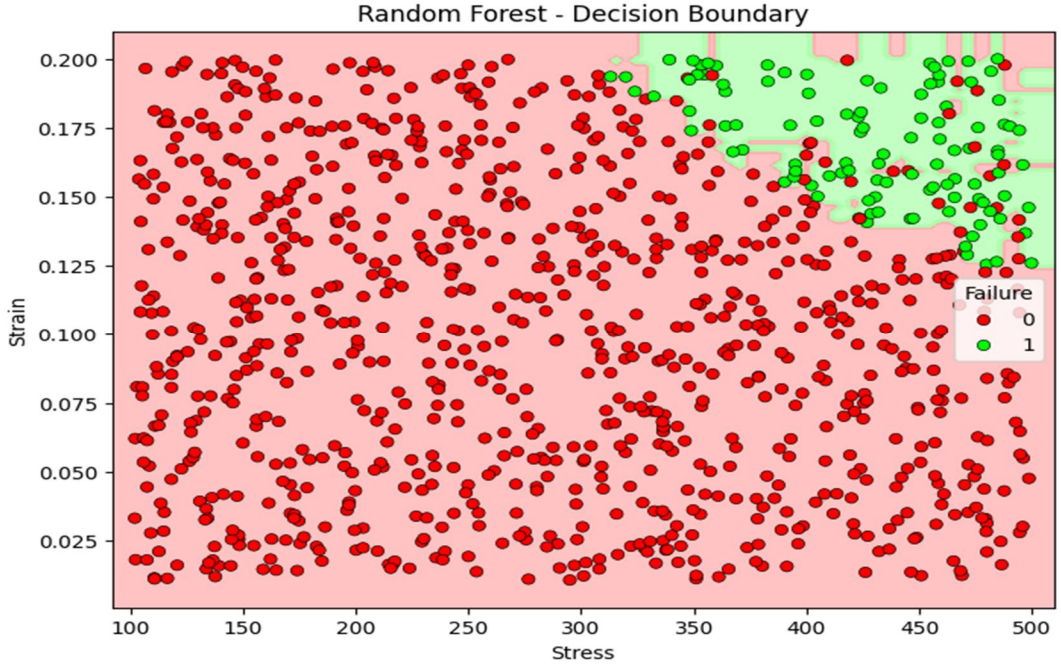


Figure 2. Decision boundary of random forest algorithm

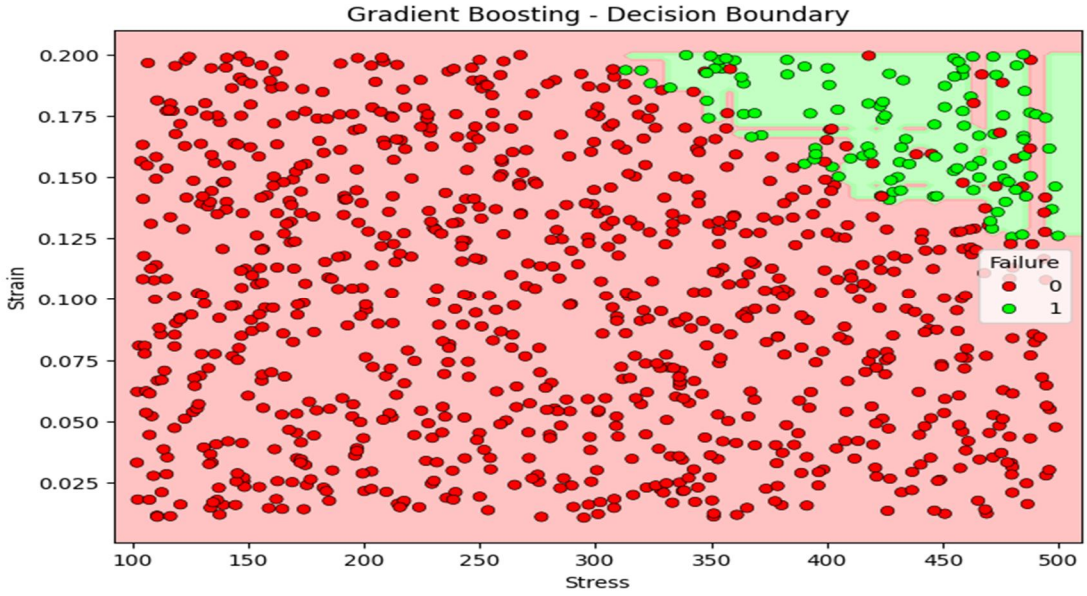


Figure 3. Decision boundary of gradient boosting algorithm

IV. CONCLUSION

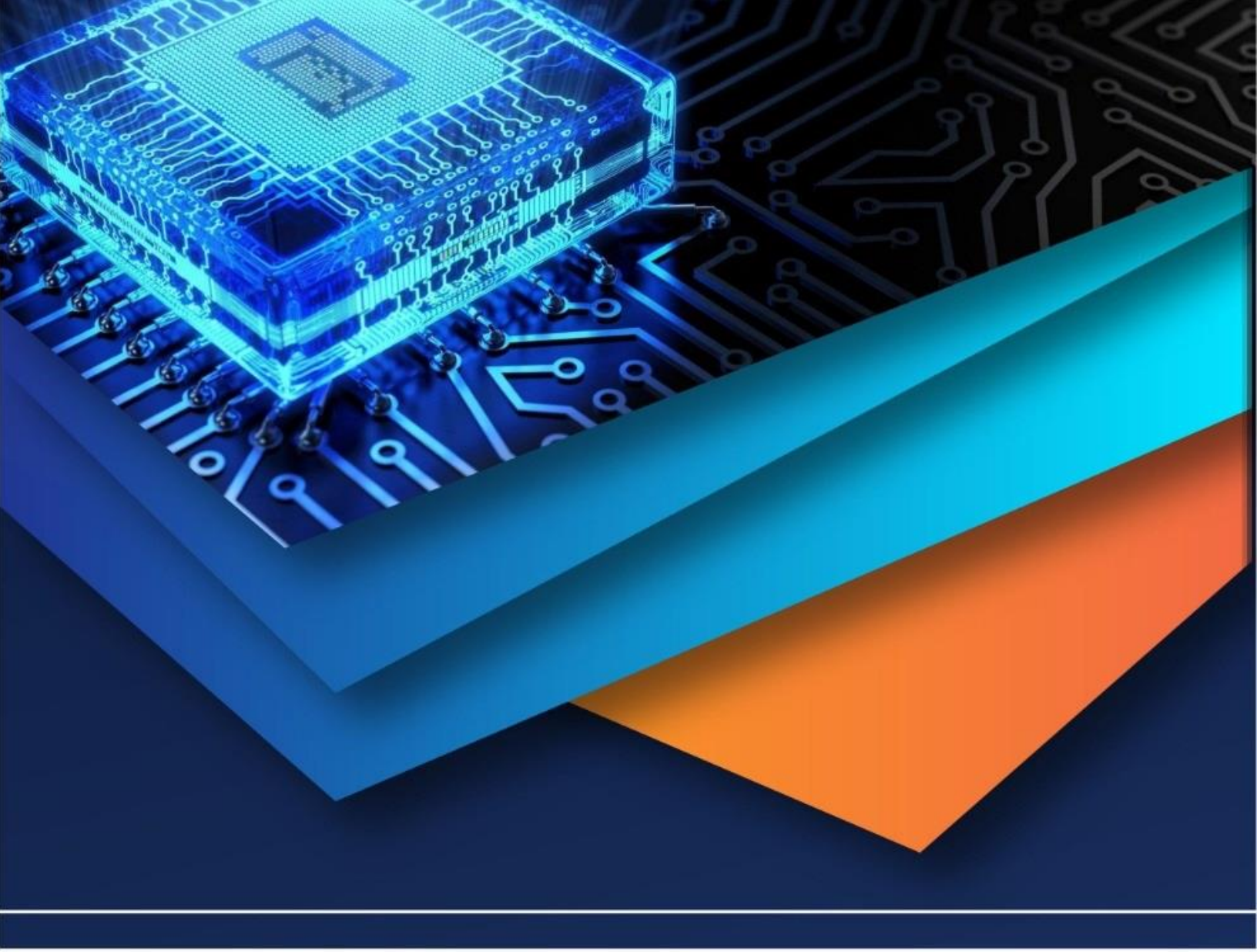
In conclusion, this study compared the performance of Random Forest and Gradient Boosting models for predicting material failure based on stress and strain data. Both models achieved strong classification results, with Gradient Boosting showing a slight advantage in recall for failure instances. The decision boundary visualization highlights that the Random Forest model effectively separates failure and non-failure regions, though some overlap and misclassification remain. The confusion matrices and classification reports confirm that both models exhibit high precision and recall, with Gradient Boosting minimizing false negatives and offering a more refined boundary.

For future work, several directions could further enhance model performance and applicability. Expanding the feature set to include additional relevant variables, such as material composition, temperature, or process parameters, could provide a more comprehensive picture of factors influencing failure. Exploring advanced algorithms, like XGBoost or stacking models, might improve precision and recall for the failure class and help reduce overfitting. Hyperparameter optimization through Bayesian Optimization or Genetic Algorithms could also yield performance gains by fine-tuning model stability and accuracy, especially in distinguishing failure cases.

REFERENCES

[1] Schwarzer, M., Rogan, B., Ruan, Y., Song, Z., Lee, D.Y., Percus, A.G., Chau, V.T., Moore, B.A., Rougier, E., Viswanathan, H.S. and Srinivasan, G., 2019. Learning to fail: Predicting fracture evolution in brittle material models using recurrent graph convolutional neural networks. Computational Materials Science, 162, pp.322-332.
[2] Zhou, S., 1998. The numerical prediction of material failure based on the material point method. The University of New Mexico.
[3] Gray, G.T., Maudlin, P.J., Hull, L.M., Zuo, Q.K. and Chen, S.R., 2005. Predicting material strength, damage, and fracture the synergy between experiment and modeling. Journal of Failure Analysis and Prevention, 5, pp.7-17.
[4] Rashid, J., Dunham, R., Rashid, M. and Machiels, A., 2009. A new material constitutive model for predicting cladding failure.
[5] Pickett, A.K., Pyttel, T., Payen, F., Lauro, F., Petrinic, N., Werner, H. and Christlein, J., 2004. Failure prediction for advanced crashworthiness of transportation vehicles. International Journal of Impact Engineering, 30(7), pp.853-872.
[6] Mazumder, R.K., Salman, A.M. and Li, Y., 2021. Failure risk analysis of pipelines using data-driven machine learning algorithms. Structural safety, 89, p.102047.
[7] Hidalgo-Mompeán, F., Fernandez, J.F.G., Cerruela-Garcia, G. and Marquez, A.C., 2021. Dimensionality analysis in machine learning failure detection models. A case study with LNG compressors. Computers in Industry, 128, p.103434.
[8] Yan, W., Deng, L., Zhang, F., Li, T. and Li, S., 2019. Probabilistic machine learning approach to bridge fatigue failure analysis due to vehicular overloading. Engineering Structures, 193, pp.91-99.
[9] Mohammed, B., Awan, I., Ugail, H. and Younas, M., 2019. Failure prediction using machine learning in a virtualised HPC system and application. Cluster Computing, 22, pp.471-485.
[10] Manikanta, J.E., Ambhore, N., Dhumal, A., Gurajala, N.K. and Narkhede, G., 2024. Machine Learning and Artificial Intelligence Supported Machining: A Review and Insights for Future Research. Journal of The Institution of Engineers (India): Series C, pp.1-11.
[11] Sana, M., Asad, M., Farooq, M.U., Anwar, S. and Talha, M., 2024. Machine learning for multi-dimensional performance optimization and predictive modelling of nanopowder-mixed electric discharge machining (EDM). The International Journal of Advanced Manufacturing Technology, 130(11), pp.5641-5664.
[12] Sarker, B., Chakraborty, S., Čep, R. and Kalita, K., 2024. Development of optimized ensemble machine learning-based prediction models for wire electrical discharge machining processes. Scientific Reports, 14(1), p.23299.
[13] Suthahar, P., Palanikumar, K., Ponshanmugakumar, A. and Anbuchezhiyan, G., 2024, October. Machine Learning Advancements in Machining Processes: A Comprehensive Review for Manufacturing Optimization. In Journal of Physics: Conference Series (Vol. 2837, No. 1, p. 012102). IOP Publishing.
[14] Hussain, A., Janjua, T.A.M., Malik, A.N., Najib, A. and Khan, S.A., 2024. Health monitoring of CNC machining processes using machine learning and wavelet packet transform. Mechanical Systems and Signal Processing, 212, p.111326.
[15] Sanga, S.S. and Antala, K.S., 2024. Cost optimization and machine learning-based prediction of waiting time for fault-tolerance machining system with general vacation and F-policy. Applied Soft Computing, 154, p.111404.
[16] Jin, L., Zhai, X., Wang, K., Zhang, K., Wu, D., Nazir, A., Jiang, J. and Liao, W.H., 2024. Big data, machine learning, and digital twin assisted additive manufacturing: A review. Materials & Design, p.113086.
[17] Mukherjee, T., Shen, W., Liao, Y. and Li, B., 2024. Improving Deposited Surface Quality in Additive Manufacturing Using Structured Light Scanning Characterization and Mechanistic Modeling. Journal of Manufacturing and Materials Processing, 8(3), p.124.
[18] Sinha, S. and Mukherjee, T., 2024. Mitigation of Gas Porosity in Additive Manufacturing Using Experimental Data Analysis and Mechanistic Modeling. Materials, 17(7), p.1569.
[19] Mishra, A., Jatti, V.S. and Paliwal, S., 2023. Evolutionary AI-Based Algorithms for the Optimization of the Tensile Strength of Additively Manufactured Specimens. In Machine Intelligence for Smart Applications: Opportunities and Risks (pp. 195-211). Cham: Springer Nature Switzerland.
[20] Mishra, A., Jatti, V.S., Sefene, E.M. and Paliwal, S., 2023. Explainable Artificial intelligence (XAI) and supervised machine learning-based algorithms for prediction of surface roughness of additively manufactured polylactic acid (PLA) specimens. Applied Mechanics, 4(2), pp.668-698.
[21] Du, Y., Mukherjee, T. and DebRoy, T., 2021. Physics-informed machine learning and mechanistic modeling of additive manufacturing to reduce defects. Applied Materials Today, 24, p.101123.
[22] Jiang, M., Mukherjee, T., Du, Y. and DebRoy, T., 2022. Superior printed parts using history and augmented machine learning. npj Computational Materials, 8(1), p.184.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)