Journal of Computer Technology and Applied Mathematics
Journal Home: https://www.suaspress.org/ojs/index.php/JCTAM | CODEN: JCTAAD
Vol. 3, No. 2, 2026 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

Research on Machine Learning–Based Prediction of Heterogeneous Metal Joining Performance and Its Application in Production and Operations Management

LI, Zhuoxuan [1*]

[1] Suzhou University of Science and Technology, CN

*** LI, Zhuoxuan is the corresponding author, E-mail: Coriander041114@outlook.com***

Abstract: Dissimilar metal joining technology is a key process for achieving lightweight structures, and accurate prediction of welding quality is crucial for ensuring structural safety. This study constructs a machine learning framework for predicting void defects in friction stir welding (FSW). Using the FSW process dataset (108 records covering three aluminum alloys: AA2219, AA2024, and AA6061), a heat input index is introduced as a derived feature, and SMOTE is applied to address class imbalance. Seven machine learning models are compared under repeated stratified five-fold cross-validation. The results show that MLP achieves the best AUC value (0.8951), followed closely by XGBoost with 0.8912 and stronger stability. This paper further explores the application of the prediction model in quality control and process optimization in a smart manufacturing environment, providing theoretical and practical references for intelligent decision-making in the welding process.

Keywords: Dissimilar Metal Joining, Friction Stir Welding, Machine Learning, Defect Prediction, Production Operations Management.

Disciplines: Artificial Intelligence. **Subjects:** Machine Learning.

DOI: https://doi.org/10.70393/6a6374616d.343037 **ARK:** https://n2t.net/ark:/40704/JCTAM.v3n2a03

1 RESEARCH BACKGROUND AND PROBLEM STATEMENT

1.1 RESEARCH BACKGROUND AND MOTIVATION

The global manufacturing industry is undergoing a profound transformation characterized by intelligence and greening. Driven by the "dual carbon" goal and energy conservation and emission reduction policies, lightweight design has become a key technological path in the fields of automobiles, aerospace, and rail transportation [1]. Dissimilar metal joining technology—especially high-quality joining of aluminum/steel, aluminum/copper, etc.—is one of the core processes for achieving structural lightweighting. By effectively joining high-strength steel with lightweight aluminum alloys, the overall weight can be significantly reduced while ensuring the structural load-bearing capacity, thereby improving energy efficiency and reducing carbon emissions. However, due to the significant differences in physical and chemical properties, dissimilar metals are prone to generating brittle intermetallic compounds (IMCs) at the interface when using traditional fusion welding methods, which seriously weakens the mechanical properties of the joint [2]. Friction stir welding (FSW), as a solid-state joining technology, has a welding temperature lower than the melting point of the material, which can effectively suppress the excessive growth of IMCs and has become the preferred solution for dissimilar metal joining [9]. Despite this, the FSW process may still produce quality problems such as voids and tunnel-type defects due to improper setting of process parameters, which may lead to catastrophic structural failure in severe cases. In the context of Industry 4.0 and intelligent manufacturing, data-driven quality prediction and process optimization are reshaping the traditional production and operation management model [3]. Machine learning technology, with its powerful nonlinear modeling capabilities, provides a new technical approach for real-time prediction and proactive intervention of welding quality. Deep integration of machine learning methods with welding processes can not only improve the quality level of a single process, but also provide data support for operational management decisions such as production planning, resource scheduling, and quality traceability.

1.2 PROBLEM STATEMENT

Current quality control of dissimilar metal FSW faces the following challenges: First, the welding process involves complex thermo-mechanical-fluid coupling behavior, and the relationship between process parameters and defect formation is highly nonlinear, which is difficult to accurately describe using traditional empirical formulas; second, the cost of acquiring experimental data is high, and the sample size available for modeling is often limited, restricting the training and verification of complex models; third, existing research focuses mainly on improving the performance of the prediction model itself, while the discussion on how the model can be embedded in the production and operation management process and support management decisions is relatively insufficient.

1.3 RESEARCH OBJECTIVES

To address the aforementioned issues, this study sets the following objectives: (1) To construct a machine learning model framework for predicting void defects in FSW (Free-Switch Welded Swing) data, and to systematically compare the prediction performance and stability of various mainstream algorithms; (2) To explore feature engineering and data augmentation strategies in small-sample data scenarios to improve the generalization ability of the model; (3) To explore the application path of machine learning prediction models in production and operation management, providing a reference for quality decision-making in the context of intelligent manufacturing.

1.4 RESEARCH CONTRIBUTIONS AND ARTICLE STRUCTURE

The main contributions of this study include: (1) Proposing a machine learning modeling process suitable for small-sample FSW data, covering feature engineering, class balancing, and robust validation strategies; (2) Through a systematic comparison of seven models, clarifying the advantages and applicable conditions of MLP and XGBoost in this task; (3) From the perspective of production and operation management, constructing a quality prediction-driven decision-making framework, expanding the application boundaries of machine learning in the field of manufacturing management.

The article is structured as follows: Chapter 2 reviews the current research status of dissimilar metal joining technology, machine learning welding quality prediction, and intelligent manufacturing operation management; Chapter 3 details the research methods, including data preprocessing, feature engineering, model building, and validation strategies; Chapter 4 reports the experimental results and provides in-depth analysis; Chapter 5 discusses the application of the model in production operation management; and finally, the article summarizes the entire paper and looks forward to future research directions.

2 CURRENT STATUS OF RELATED RESEARCH

2.1 DISSIMILAR METAL JOINING TECHNOLOGY

The core challenge of dissimilar metal joining lies in overcoming the differences in physicochemical properties between different materials. Taking aluminum/steel joining as an example, the melting points of the two metals differ by about 800°C, and their coefficients of thermal expansion differ by more than two times. Furthermore, aluminum readily reacts with iron at high temperatures to form Fe-Al intermetallic compounds [2]. Studies have shown that when the IMC layer thickness exceeds 10 μm, the joint strength will decrease sharply [4].

Traditional fusion welding methods (such as MIG and TIG welding) have excessively high peak temperatures, making it difficult to effectively control IMC growth, and the joint performance often fails to meet engineering requirements. Solid-state joining technology exhibits unique advantages due to its welding temperature being lower than the material melting point. Friction stir welding (FSW), invented by the Welding Institute (TWI) in the UK in 1991, has been widely used in joining aluminum alloys, magnesium alloys, and dissimilar metals [9]. FSW achieves solid-state mixing of materials through frictional heat and plastic deformation between the high-speed rotating stirring head and the workpiece. The IMC layer thickness can be controlled within a few micrometers, and the joint strength is significantly better than that of fusion welding methods.

However, FSW is extremely sensitive to process parameters. Even small changes in parameters such as rotational speed, welding speed, axial pressure, and stirring head geometry can lead to defects. Void defects are one of the most common volumetric defects in FSW, mainly caused by insufficient material flow or uneven plastic deformation [14]. The presence of voids significantly weakens the fatigue life and static strength of the joint, therefore their prediction and control have significant engineering value.

2.2 APPLICATION OF MACHINE LEARNING IN WELDING QUALITY PREDICTION

Machine learning has made significant progress in the field of welding quality prediction in recent years. Early studies mostly used artificial neural networks (ANNs) to establish the mapping relationship between process parameters and weld quality [5]. With the development of algorithms, support vector machines, random forests, gradient boosting trees, and other methods have been successively introduced into welding quality prediction tasks [15, 17].

Ensemble learning methods have received widespread attention due to their excellent generalization performance. Breiman's random forest effectively reduces the risk of overfitting by integrating multiple decision trees [15]. XGBoost introduces a regularization term in the gradient boosting framework, significantly improving the robustness and training efficiency of the model [17]. In recent years, new-generation gradient boosting algorithms such as LightGBM and CatBoost have demonstrated stronger capabilities in handling large-scale data and class features [18, 19].

Deep learning methods have also been applied to welding defect detection. Convolutional neural networks (CNNs) can automatically extract defect features from weld X-ray images or ultrasonic signals, achieving end-to-end defect identification [6]. However, deep learning requires a large amount of data, limiting its application in small-sample industrial scenarios.

To address the common class imbalance problem in welding data, oversampling techniques such as SMOTE have been widely adopted [11]. Studies have shown that reasonable data augmentation strategies can significantly improve the accuracy of minority class identification, which is particularly crucial for tasks like defect detection where positive and negative samples are severely imbalanced.

2.3 PRODUCTION OPERATIONS MANAGEMENT IN SMART MANUFACTURING

Under the Industry 4.0 framework, production operations management is transitioning from experience-driven to data-driven [3]. Smart manufacturing emphasizes the deep integration of physical and information systems, achieving self-perception, self-decision-making, and self-optimization of the production process through real-time data acquisition, analysis, and feedback. Digital twin technology provides a new paradigm for the virtual mapping and simulation optimization of welding processes [7]. By constructing a digital twin of the welding process, process parameters can be rapidly iterated and optimized in a virtual environment, reducing the cost of physical testing. Machine learning models, as the core algorithm component of digital twins, can predict welding quality under different parameter combinations based on historical data, providing a quantitative basis for process decision-making. Predictive maintenance is another important application scenario of intelligent operation management [8]. By monitoring the operating status parameters of welding equipment and combining them with machine learning models, equipment failure risks can be predicted in advance, and maintenance plans can be reasonably arranged, thereby improving overall equipment efficiency (OEE) and reducing unplanned downtime losses.

2.4 RESEARCH GAPS

Based on the above literature analysis, the current research has the following gaps: First, comparative studies of machine learning for predicting void defects in FSW are relatively limited, especially regarding the applicability of various algorithms in small sample data scenarios, which requires further in-depth exploration; Second, the role of feature engineering in welding quality prediction has not received sufficient attention, and feature construction strategies based on physical mechanisms need systematic research; Third, existing research mainly focuses on the technical performance of prediction models, while there is little discussion on how models can be integrated into production operation management processes and support management decisions. This study aims to fill these gaps, construct a machine learning framework for FSW defect prediction, and explore its application path in intelligent manufacturing operation management.

3 METHODOLOGY

3.1 RESEARCH FRAMEWORK

This research aims to construct a machine learning framework for predicting the quality of friction stir welding of dissimilar metals, enabling accurate identification and early warning of void defects during the welding process. The framework comprises four core components: data acquisition and preprocessing, feature engineering, model building and training, and cross-validation evaluation. The overall technical approach follows the standard paradigm of data-driven modeling, while also being specifically optimized for small-sample data scenarios.

The experimental platform is based on a computing environment equipped with an NVIDIA GeForce RTX 3050 graphics card (8GB VRAM), using CUDA 13.1 for GPU acceleration. The programming environment is Python 3.11, with key dependencies including scikit-learn, XGBoost, LightGBM, CatBoost, and imbalanced-learn.

3.2 DATA DESCRIPTION AND PREPROCESSING

3.2.1 Data Sources

This study uses the "Void Formation Process Data in Welding" dataset publicly available on the Kaggle platform. This dataset originates from friction stir welding (FSW) experiments and records the formation of void defects in welded joints under different combinations of process parameters. FSW, as a solid-state joining technology, has attracted widespread attention due to its ability to effectively suppress the formation of brittle intermetallic compounds in dissimilar metal joining [9]. However, improper process parameter settings can still lead to quality problems such as voids and tunnel-type defects, which can cause catastrophic structural failure in severe cases.

The dataset contains 108 experimental records, covering FSW experimental data for three aluminum alloys: AA2219, AA2024, and AA6061. The original features include 11 process parameters and material property variables: Alloy type, Welding speed, Rotation speed, Plate thickness, Shoulder radius, Axial pressure, Pin root radius, Pin tip radius, Tilt angle, Thermal diffusivity, and Yield strength. The label variables are binary classification targets, where "1" indicates the presence of void defects and "0" indicates that the weld quality is qualified.

3.2.2 Data preprocessing

The original data has some outlier labels (such as "*" symbols after the numerical values) and a small number of missing values. The preprocessing process first cleans the string data, removes special characters and converts it to numerical type; then, the median imputation strategy is used to handle missing values. This method is more robust to outliers than mean imputation [10]. The label distribution statistics show that there are 65 defect-free samples and 43 defective samples in the dataset, with a class ratio of approximately 1.5:1, indicating a certain degree of class imbalance. To alleviate the impact of class imbalance on model training, this study introduces the SMOTE (Synthetic Minority Over-sampling Technique) oversampling method [11]. SMOTE generates synthetic samples by linear interpolation between minority class samples, which can effectively expand the minority class size without introducing the risk of overfitting caused by duplicate data. It should be emphasized that SMOTE is only applied within the training set, and the validation set always maintains the original distribution to ensure the objectivity of the evaluation results. In terms of feature standardization, the Z-score standardization method is used to transform each feature to a standard normal distribution with a mean of 0 and a standard deviation of 1. The standardization parameters (mean and standard deviation) are only calculated from the training set and used to transform the validation set, thereby strictly avoiding the problem of data leakage [12].

3.3 FEATURE ENGINEERING

In addition to the original 11 features, this study attempts to construct several derived features based on the welding physical mechanism to enhance the model's representation ability. The derived features initially explored include: stirring pin geometry ratio (Pin_Geometry_Ratio = Pin_root_radius / Pin_tip_radius), energy density index (Energy_Density = Rotation_speed × Axial_pressure / Welding_speed), plate thickness-shoulder ratio (Thickness_Shoulder_Ratio = Plate_thickness / Shoulder_radius), and heat input index (Heat_Input_Index = Rotation_speed / Welding_speed).

Comparative experiments revealed that introducing multiple derived features simultaneously led to a decrease in model performance and a significant increase in variance. This phenomenon can be attributed to the curse of dimensionality effect on small sample datasets [13]. After testing and screening one by one, only Heat_Input_Index was ultimately retained as a new feature. This feature has a clear physical meaning: the ratio of rotational speed to welding speed directly reflects the amount of heat input per unit stroke, and heat input is a key factor controlling the material flow behavior and defect formation in the FSW process [14]. After adding this feature, the model's predictive performance was steadily improved without introducing additional instability.

The final feature set contains 12 variables, which are used for the training and evaluation of subsequent machine learning models.

3.4 MACHINE LEARNING MODELS

This study selected seven representative machine learning algorithms for comparative experiments, covering three major categories: ensemble learning methods, support vector machines, and neural networks.

Random Forest, proposed by Breiman in 2001, constructs multiple decision trees through bootstrap sampling and uses a voting mechanism for prediction, exhibiting good resistance to overfitting [15]. In this study, the forest size was set to 100 trees, with a maximum depth limit of 5 layers to control model complexity.

Gradient Boosting Decision Tree (GBDT) adopts a forward stepwise addition model, iteratively fitting the prediction residuals of the previous round to gradually optimize model performance [16]. XGBoost introduces regularization terms and second-order Taylor expansion optimization on the basis of GBDT, significantly improving training efficiency and generalization ability [17]. LightGBM adopts gradient-based one-sided sampling (GOSS) and mutually exclusive feature binding (EFB) strategies, significantly reducing computational overhead while maintaining accuracy [18]. CatBoost proposes an ordered target statistics encoding method for class features, effectively avoiding the target leakage problem introduced by traditional encoding methods [19]. All four gradient boosting algorithms mentioned above are set to 100 iterations, a maximum depth of 4, and a learning rate of 0.1.

Support Vector Machines (SVMs) classify samples by finding the maximum margin hyperplane [20]. This study uses the Radial Basis Function (RBF) kernel, with a regularization coefficient C set to 1.0, and an adaptive strategy for the kernel parameter gamma.

Multi-Layer Perceptrons (MLPs) are a typical example of feedforward neural networks, learning complex mapping relationships between inputs and outputs through multi-layer nonlinear transformations [21]. The MLP constructed in this study contains two hidden layers with 64 and 32 neurons respectively, uses the ReLU activation function, and has a maximum iteration count of 800.

All models have a fixed random seed of 66 to ensure reproducibility of experimental results.

3.5 MODEL EVALUATION METRICS

Given the small size of the dataset (only 108 records), traditional training-test set splitting methods are insufficient to obtain robust performance estimates. This study employs a repeated stratified 5-Fold Cross-Validation strategy, repeated 10 times for a total of 50 independent evaluations [22]. Stratified sampling ensures that the ratio of positive to negative samples in each fold remains consistent with the original dataset, while multiple repetitions effectively reduce the estimation variance caused by random splitting.

Model performance is evaluated using three core metrics:

(1) AUC (Area Under the ROC Curve): The area under the ROC curve measures the overall discriminative power of the classifier under different threshold settings, ranging from 0.5 to 1.0. A higher value indicates a stronger ability to rank positive and negative samples [23]. AUC is threshold-independent and provides a more comprehensive performance evaluation than a single accuracy rate.

(2) Accuracy: The proportion of correctly classified samples to the total number of samples is the most intuitive classification performance metric.

(3) F1 Score: The harmonic means of precision and recall, which is more valuable than accuracy in class imbalance scenarios [24]. Considering both predictive performance and practical application needs, this study uses AUC as the primary basis for model selection, with F1 score and accuracy as auxiliary references.

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 EXPERIMENTAL SETUP

The experiments strictly followed the methodological framework described in Chapter 3. During data preprocessing, missing value imputation, outlier cleaning, and feature standardization were performed. SMOTE oversampling was applied independently to the training set for each fold cross-validation, while the validation set maintained its original distribution. Seven machine learning models were trained and evaluated under the same data partitioning. Finally, the results of 50 rounds of cross-validation were summarized to calculate the mean and standard deviation of each metric.

4.2 MODEL PERFORMANCE COMPARISON

Table 1 summarizes the performance of the seven models under repeated five-fold cross-validation, with results arranged in descending order of AUC.

TABLE 1: CROSS-VALIDATION PERFORMANCE COMPARISON OF EACH MODEL (MEAN ± STANDARD DEVIATION)

Model	AUC (mean ± std)	Accuracy (mean ± std)	F1
SVM (RBF)	0.7035 (± 0.1163)	0.6065 (± 0.1275)	0.5570
Random Forest	0.8324 (± 0.1018)	0.7497 (± 0.1076)	0.7033
LightGBM	0.8647 (± 0.0750)	0.7818 (± 0.0850)	0.7360
CatBoost	0.8657 (± 0.0835)	0.7689 (± 0.0924)	0.7274
Gradient Boosting	0.8724 (± 0.0848)	0.7960 (± 0.0891)	0.7406
XGBoost	0.8912 (± 0.0758)	0.8080 (± 0.0971)	0.7658
MLP (Neural Net)	**0.8951 (± 0.1005)**	**0.8503 (± 0.0874)**	**0.8096**

From the AUC metric, the MLP neural network ranks first with a mean of 0.8951, followed closely by XGBoost with a slight difference (0.8912). The performance difference between the two is not statistically significant. Gradient boosting algorithms (Gradient Boosting, CatBoost, and LightGBM) generally perform similarly, with AUC values all falling between 0.86 and 0.87. Random forest's AUC is 0.8324, slightly lower than the ensemble boosting methods. SVM performs relatively weakly, with an AUC of only 0.7035, which may be related to the difficulty of constructing an effective decision boundary in the current feature space using the RBF kernel function.

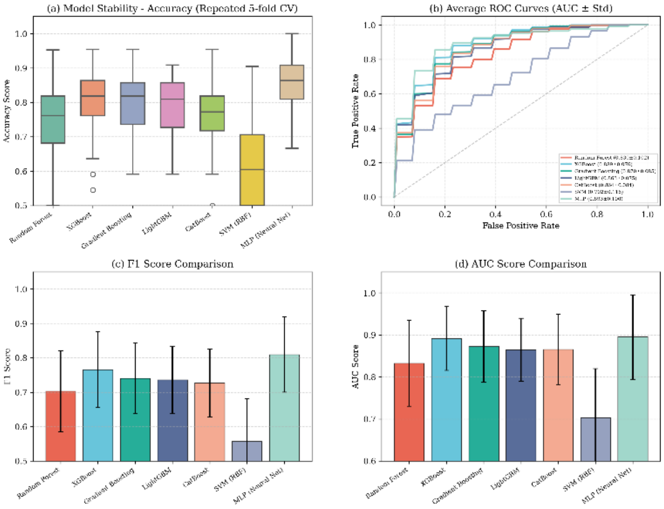

FIG. 1. OVERALL PERFORMANCE COMPARISON OF EACH MODEL

Figure 1 shows the overall performance visualization results of each model, including four subplots: (a) accuracy boxplot, (b) average ROC curve, (c) F1 score comparison, and (d) AUC score comparison.

4.3 MODEL STABILITY ANALYSIS

Figure 1(a) shows the box plots, which visually reflect the differences in stability among the models after 50 rounds of cross-validation. XGBoost and LightGBM exhibit the best stability, with relatively narrow interquartile ranges (IQRs) and fewer outliers in their accuracy distributions. While MLP has the highest mean, its box span is slightly larger than XGBoost's, indicating some performance fluctuations in small-sample scenarios. SVM shows the worst stability, with a low median and multiple outliers significantly deviating from the main distribution, suggesting severe underfitting in some data partitions.

Further analysis from the perspective of standard deviation reveals that LightGBM has the smallest standard deviation in the AUC metric (0.0750), demonstrating the strongest generalization stability. In contrast, the standard deviations of SVM and MLP both exceed 0.10, reflecting higher uncertainty. Considering both mean and variance, XGBoost achieves a good balance between predictive performance and stability.

4.4 ROC CURVE ANALYSIS

Figure 1(b) shows the average ROC curves of the seven models. It can be observed that the curves of MLP and XGBoost are closest to the upper left corner, maintaining a high true positive rate throughout the entire false positive rate range. Notably, in the low false positive rate region (FPR < 0.2), XGBoost performs slightly better than MLP, which is significant for applications requiring strict control of false positive rates in actual production.

The curves of Gradient Boosting, LightGBM, and CatBoost highly overlap, confirming the conclusion in Table 1 that their AUC values are similar. The ROC curve of Random Forest is generally below the gradient boosting series, but still significantly better than the diagonal reference line, indicating that it has a certain classification ability. The ROC curve of SVM shows a significant dip in the middle region, reflecting the model's insufficient discrimination ability in a specific threshold range.

4.5 DISCUSSION

Experimental results show that the machine learning-based welding defect prediction method can effectively identify the risk of void formation in the FSW process. Both MLP and XGBoost achieved an AUC of nearly 0.90 on this dataset, validating the applicability of deep learning and ensemble learning methods for this task.

The optimal AUC achieved by the neural network (MLP) can be attributed to its powerful nonlinear fitting ability. The relationship between process parameters and defect formation during welding often exhibits highly nonlinear characteristics, and MLP can capture this complex mapping through cascaded transformations of multiple hidden units. However, the high standard deviation of MLP also suggests the inherent limitations of neural networks on small sample data, namely, their susceptibility to fluctuations in training data.

XGBoost's robust performance is attributed to its built-in regularization mechanism and native support for missing values. Given the prevalence of noise and incompleteness in engineering data, XGBoost demonstrates stronger robustness. Furthermore, the gradient boosting method offers better interpretability than neural networks, making it easier for engineers to understand the model's decision-making process.

SVM performed poorly in this experiment, possibly due to the following reasons: First, the RBF kernel function is sensitive to hyperparameters (C and gamma), and the default parameter settings used in this study may not be optimal; second, 108 samples are insufficient for SVM, a method that requires sufficient support vectors; third, the standardized feature space may still contain linearly inseparable regions, limiting the classification performance of SVM.

The effectiveness of the `Heat_Input_Index` feature validates the important role of domain knowledge in feature engineering. Heat input, as a core parameter controlling the flow of FSW material, effectively captures the interaction effect between welding speed and rotation speed through its ratio expression. The poor performance of other derived features may be due to the introduction of redundant information or high collinearity with the original features, which exacerbates the risk of overfitting under small sample conditions.

It should be noted that this study still has several limitations. First, the data size of 108 samples limits the training potential of complex models and restricts the exploration of more derived features. Second, the dataset only covers three aluminum alloy materials, and the transferability of the model to other dissimilar metal combinations needs to be verified. Finally, the real-time deployment and online learning mechanisms of the model have not yet been addressed, which will be an important direction for future research.

5 APPLICATION IN PRODUCTION AND OPERATIONS MANAGEMENT

5.1 QUALITY PREDICTION-DRIVEN DECISION FRAMEWORK

Based on the experimental conclusions in Chapter 4, this section constructs a decision framework that embeds a machine learning prediction model into the production operations management process. This framework, with quality prediction at its core, connects the three stages of process design, production execution, and quality control, achieving data-driven closed-loop management.

In the process design stage, the prediction model can serve as a virtual testing tool to assist engineers in parameter selection. Traditional process development relies on numerous physical experiments, which is time-consuming and costly. With a trained MLP or XGBoost model, engineers can quickly assess the defect risk of different parameter combinations, select high-potential solutions, and then conduct a limited number of physical verifications, thereby significantly shortening the process development cycle.

In the production execution stage, the prediction model can be integrated with an online monitoring system to achieve real-time early warning of welding quality. When the process parameters collected by sensors are input into the model, if the predicted defect probability exceeds a preset threshold, the system can automatically trigger an alarm or adjustment command to avoid batch quality accidents. The XGBoost model in this study combines high AUC value and stability, making it suitable for deployment in online prediction scenarios with high reliability requirements.

In the quality control stage, the output of the prediction model can be used as the basis for formulating the sampling inspection strategy. For welds predicted by the model to be of high risk, the sampling ratio of non-destructive testing can be increased; while for low-risk welds, the inspection frequency can be appropriately reduced, thereby optimizing the allocation of inspection resources while ensuring quality.

5.2 PROCESS PARAMETER OPTIMIZATION

Machine learning models can not only be used for quality prediction, but also as the objective function of optimization algorithms to achieve automatic optimization of process parameters. Specifically, the trained prediction model can be combined with metaheuristic algorithms such as genetic algorithms and particle swarm optimization: with the minimization of defect probability as the optimization objective and the feasible region of process parameters as the constraint, the optimal or near-optimal parameter combination can be searched [25]. Taking the FSW dataset of this study as an example, the heat input index (Heat_Input_Index = Rotation_speed / Welding_speed) is identified as a key derived feature affecting void defects. In production practice, this index can be used as a core control parameter for the process window: when the heat input is too low, the material is not sufficiently plasticized, has poor fluidity, and is prone to voids; when the heat input is too high, it may lead to excessive softening or even local melting of the material. Determining the reasonable range of the heat input index through model prediction helps guide the fine adjustment of process parameters on-site.

Furthermore, the feature importance assessment function built into gradient boosting models (such as XGBoost) can provide directional guidance for process optimization. Engineers can prioritize parameters with high importance rankings, concentrate resources on in-depth control, and adopt standardized settings for secondary parameters, thereby improving optimization efficiency.

5.3 MANAGEMENT IMPLICATIONS

From the perspective of production operation management, the findings of this study have the following implications for manufacturing enterprises:

First, the cultivation of data asset awareness. The publicly available dataset used in this study contains only 108 records, yet it is sufficient to support the training of a prediction model with an AUC close to 0.90. This indicates that even small- to medium-sized historical data has significant analytical value after standardization. Manufacturing enterprises should attach importance to the systematic collection and storage of welding process data to lay the foundation for subsequent data analysis and modeling. Second, the importance of cross-domain knowledge integration. The successful construction of the Heat_Input_Index feature in this study stems from the organic combination of welding process domain knowledge and machine learning methods. Purely data-driven feature selection has limited effectiveness in small sample scenarios, while feature engineering incorporating domain mechanisms can effectively improve the physical interpretability and generalization ability of the model. This suggests that enterprises should promote deep collaboration between process experts and data scientists when advancing intelligent transformation.

Third, the trade-offs in model selection. Experimental results show that MLP achieved the highest mean AUC, but XGBoost has advantages in stability and interpretability. In practical deployment, enterprises need to make trade-offs based on the specific needs of the application scenario: for online prediction systems requiring high reliability, stability may be more important than peak performance; while for offline analysis or batch evaluation tasks, prediction accuracy can be prioritized.

6 CONCLUSION

This study addresses the quality prediction problem in friction stir welding of dissimilar metals, constructing a systematic machine learning modeling framework and exploring its application potential in production operation management. The study utilizes the publicly available FSW void defect dataset from the Kaggle platform, employing SMOTE oversampling to address class imbalance and proposing a thermal input index as an effective derived feature based on welding physics. Seven mainstream machine learning algorithms are systematically compared and evaluated using a repeated hierarchical five-fold cross-validation framework. Experimental results show that the MLP neural network achieves the best prediction performance with an AUC of 0.8951, followed closely by XGBoost with an AUC of 0.8912 and exhibiting stronger stability, while SVM performs relatively weakly on this task. The study further explores the path and value of integrating machine learning models into intelligent manufacturing operation management processes from three dimensions: quality prediction-driven decision-making, process parameter optimization, and management implications.

This study has several limitations: the relatively small dataset size limits the training potential of complex models; the experiments only cover aluminum alloy materials, and the model's transferability to dissimilar metal combinations such as aluminum/steel needs further verification; in addition, real-time deployment and online incremental learning mechanisms of the model are not yet addressed. Future research can be conducted in the following directions: expanding the scale of data and incorporating more material combinations and defect types; introducing interpretable artificial intelligence (XAI) methods to deeply analyze the model's decision-making mechanism; exploring the application of deep learning and transfer learning in cross-process and cross-equipment scenarios; and carrying out deployment verification and effect evaluation of the model on real production lines.

ACKNOWLEDGMENTS

Not Applicable.

FUNDING

Not Applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not Applicable.

INFORMED CONSENT STATEMENT

Not Applicable.

DATA AVAILABILITY STATEMENT

Not Applicable.

CONFLICT OF INTEREST

Not Applicable.

PUBLISHER'S NOTE

AUTHOR CONTRIBUTIONS

Not application.

ABOUT THE AUTHORS

LI, Zhuoxuan

Suzhou University of Science and Technology, CN, Coriander041114@outlook.com.

REFERENCES

[1] Taub, A., De Moor, E., Luo, A., Matlock, D. K., Speer, J. G., & Vaidya, U. (2019). Materials for automotive lightweighting. Annual Review of Materials Research, 49(1), 327-359.

[2] Springer, H., Kostka, A., Payton, E. J., Raabe, D., Kaysser-Pyzalla, A., & Eggeler, G. J. A. M. (2011). On the formation and growth of intermetallic phases during interdiffusion between low-carbon steel and aluminum alloys. Acta materialia, 59(4), 1586-1600.

[3] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. Journal of manufacturing systems, 48, 157-169.

[4] Shahani, A. R., & Farrahi, A. (2019). Experimental investigation and numerical modeling of the fatigue crack growth in friction stir spot welding of lap-shear specimen. International Journal of Fatigue, 125, 520-529.

[5] Pal, S., Pal, S. K., & Samantaray, A. K. (2008). Artificial neural network modeling of weld joint strength prediction of a pulsed metal inert gas welding process using arc signals. Journal of materials processing technology, 202(1-3), 464-474.

[6] Zhang, Y., You, D., Gao, X., Zhang, N., & Gao, P. P. (2019). Welding defects detection based on deep learning with multiple optical sensors during disk laser welding of thick plates. Journal of Manufacturing Systems, 51, 87-94.

[7] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. (2018). Digital twin in industry: State-of-the-art. IEEE Transactions on industrial informatics, 15(4), 2405-2415.

[8] Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024.

[9] Mishra, R. S., & Ma, D. Z. (2005). Friction stir welding and processing. Materials science and engineering: R: reports, 50(1-2), 1-78.

[10] Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

[11] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[12] Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. (No Title).

[13] Bellman, R. (1961). A mathematical formulation of variational processes of adaptive type. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Vol. 4, pp. 37-49). University of California Press.

[14] Nandan, R., DebRoy, T., & Bhadeshia, H. K. D. H. (2008). Recent advances in friction-stir welding–process, weldment structure and properties. Progress in materials science, 53(6), 980-1023.

[15] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[16] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

[17] Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. Cornell University.

[18] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

[19] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.

[20] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[21] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

[22] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

[23] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), 1145-1159.

[24] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

[25] Gupta, S. K., Pandey, K. N., & Kumar, R. (2018). Multi-objective optimization of friction stir welding process parameters for joining of dissimilar AA5083/AA6063 aluminum alloys using hybrid approach. Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications, 232(4), 343-353.