APEM journal

Advances in Production Engineering & Management
Volume 21 | Number 1 | March 2026 | pp 16–24
https://doi.org/10.14743/apem2026.1.558

ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

A calibrated ensemble framework for multi-class defect evaluation in ceramic sanitaryware manufacturing

Ozcetin, E.[a,*], **Simsek, S.**[a]

[a]Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Hitit University, Corum, Turkiye

A B S T R A C T

Reliable defect evaluation is essential in ceramic sanitaryware manufacturing, where inspection outcomes directly influence rework decisions, process control, and delivery performance. In practice, defect assessment is often affected by operational variability, class imbalance, and human-dependent inspection procedures, which limit the repeatability and consistency of quality control decisions. This study formulates multi-class defect classification as a practical quality control problem and investigates the robustness of production-data-based decision-support in an industrial environment. The analysis is based on a real-world dataset comprising 11,071 production records collected under routine operating conditions. Defect labels were assigned through a two-stage quality control procedure involving trained inspectors and supervisory verification. Multinomial logistic regression, support vector machines with radial basis function kernels, and CatBoost were evaluated as base classifiers. A probability-based voting ensemble was developed to integrate the complementary decision structures, and posterior probabilities were calibrated using Platt scaling prior to aggregation to improve decision consistency. Experimental results show that the proposed calibrated ensemble improves Macro-F1 from 0.7126 (logistic regression) to 0.7211 and enhances minority-class recall, leading to more balanced performance under severe class imbalance. The findings indicate that probability calibration and ensemble integration contribute to improved stability and interpretability of defect evaluation. Overall, the proposed framework provides a practically deployable decision-support layer that supports more consistent rework decisions and more reliable quality control in industrial production settings.

A R T I C L E I N F O

Keywords:
Ceramic sanitaryware;
Defect classification;
Production data;
Class imbalance;
Ensemble learning;
Probability calibration;
Industrial quality control

**Corresponding author:*
erdenerozcetin@hitit.edu.tr
(Ozcetin, E.)

Article history:
Received 18 March 2026
Revised 6 April 2026
Accepted 13 April 2026

1. Introduction

Ceramic sanitaryware products such as washbasins, toilets, and bidets are required to meet high standards in terms of aesthetics, hygiene, and mechanical durability [1]. Accordingly, quality control plays a critical role in ceramic sanitaryware manufacturing, where production defects such as cracks, retouch-related imperfections, and shape irregularities directly translate into reprocessing costs, increased scrap rates, and delivery risks. In today's highly competitive market and large-scale production environments, accurate and consistent defect evaluation is therefore essential not only as a manufacturing activity, but also as a quality control task in which the defect state of each product is inferred from multiple sources of evidence.

In practice, inspection in ceramic sanitaryware production is still predominantly performed through manual visual assessment by experienced operators. This corresponds to a human-centered inspection procedure whose output is affected by operator-dependent variability, fatigue, and attention, as well as by changing observation conditions. Moreover, the high throughput of modern production lines limits the feasibility of fully manual inspection. These limitations motivate automated inspection frameworks that can be regarded as industrial inspection systems, where the measurand is the defect-related condition of the product surface and geometry, and the overall inspection outcome depends on the inspection context (illumination, surface texture, product positioning, and data acquisition settings) and subsequent data processing.

Early studies on ceramic defect detection primarily focused on image processing and morphological analysis techniques [2-4]. While these approaches demonstrated that several surface defects can be detected algorithmically, their robustness under real production conditions has often remained limited. In industrial production settings, defect assessment is highly influenced by operational variability, including changes in process conditions, heterogeneous manufacturing environments, and differing reporting practices. Unless evaluation procedures are deliberately designed to account for such factors, the repeatability of inspection decisions may be compromised. In line with this perspective, Karimi and Asemani [5] noted that no single inspection technique is capable of capturing the full range of ceramic defect mechanisms. In other words, these findings suggest that improving inspection practice requires not only the development of advanced recognition algorithms, but also a clearer definition of the industrial context and a systematic evaluation of system reliability under variable operating conditions.

In addition to traditional image-based inspection, researchers have increasingly turned to data-driven methods to analyze defect formation by incorporating production and process variables. Dengiz *et al*. [6] showed that data mining techniques can be used to relate defect characteristics to specific manufacturing practices, whereas Kesharaju and Nagarajah [7] reported that integrating optimization strategies with learning-based models leads to improved classification performance. From a production-oriented perspective, such approaches can be interpreted as data processing and estimation mechanisms that integrate heterogeneous operational information with inspection outcomes to support decision-making. However, to be informative in practical industrial settings, these studies should go beyond accuracy reporting and explicitly address the industrial context, evaluation procedures, and robustness of results under realistic sources of operational variability and class imbalance.

Despite these advances, a persistent challenge in ceramic sanitaryware defect classification is the imbalanced and partially overlapping structure of defect categories. Certain defects, such as cracks, are relatively frequent and comparatively separable. On the other hand, others, including retouch-related defects, are rarer and share characteristics with multiple fault mechanisms. This imbalance makes model training more challenging and can lead to biased performance if the evaluation procedure is not designed to reflect operational priorities. In practical terms, the goal is not only to improve predictive performance but also to obtain reliable and stable decision outcomes across defect types, minimizing systematic misclassification of minority yet operationally important categories.

Motivated by these considerations, this study investigates multi-class defect evaluation in ceramic sanitaryware manufacturing using a real-world industrial dataset comprising 11,071 production observations. The dataset is derived directly from operational records and is analyzed with minimal intervention to preserve authentic production characteristics. Owing to sparsity in several defect categories, the task is formulated using three industrially relevant classes: Crack, Retouch, and Other. The proposed analysis is positioned as a production-data-based quality-control and decision-support layer that operates on routinely collected production information, aiming to enhance the consistency and practical usability of inspection outcomes. Unlike prior ceramic defect studies that mainly report accuracy, this study emphasizes decision reliability by integrating probability calibration with ensemble-based evaluation under severe class imbalance using real-world production data.

We comparatively evaluate multinomial logistic regression, support vector machines with a radial basis function kernel, and CatBoost. In addition, we propose a probability-based voting ensemble that integrates the complementary strengths of individual classifiers. Particular attention is given to categorical feature handling, probability calibration, and systematic hyperparameter optimization to support a more disciplined performance evaluation under class imbalance. In detail, the experimental design emphasizes metrics that better reflect operational reliability under imbalance (e.g., class-sensitive performance measures), and reports results in a way that facilitates replication in similar industrial quality control settings.

This study makes three main contributions. First, it provides a comprehensive empirical analysis of defect evaluation based on large-scale, real production data, with an explicit focus on industrial quality control and inspection reliability in ceramic sanitaryware manufacturing. Second, it demonstrates that a carefully designed probability-based ensemble can yield more balanced and reliable performance under severe class imbalance, supporting robust decision-making at high throughput. Third, by incorporating process-related variables and relatively interpretable modeling components, the proposed framework offers a practical decision-support tool that can be integrated into existing quality control workflows in ceramic sanitaryware manufacturing.

2. Background

Automated defect detection and classification in ceramic manufacturing have received increasing attention due to the limitations of manual inspection and the increasing complexity of production processes. Early studies mostly focused on image processing and morphological analysis to identify surface defects on ceramic products. Forest [2] demonstrated that relatively simple image-based techniques could be used to identify visible surface irregularities, thereby providing an early foundation for automated inspection systems. Elbehiery *et al.* [3] subsequently introduced a low-cost and portable solution for detecting common defects, including cracks and pinholes, through morphological processing. In a similar vein, Hocenski *et al.* [4] employed image analysis methods to recognize edge defects, scratches, glazing faults, and texture anomalies, illustrating that several inspection activities could be partially automated.

Despite these developments, later studies pointed out that purely image-based inspection methods often lack robustness under real production conditions [8]. Rahaman and Hossain [9] observed that performance obtained in controlled environments tends to decline when faced with low contrast, surface noise, and fluctuating illumination. Karimi and Asemani [5] also noted that no single image processing technique can adequately represent the wide range of ceramic defect types, thereby underscoring the need for complementary approaches. In sum, these results suggest that inspection performance is strongly affected by operational variability, which in turn limits reliability and transferability across different manufacturing settings.

Parallel to classical vision-based inspection, data-driven approaches using production and process variables have been explored to support defect evaluation [6]. In a related direction, Kumru [10] proposed a fuzzy logic-based evaluation framework for sanitaryware quality assessment to reduce subjectivity and improve decision consistency. These lines of work motivate defect evaluation strategies that leverage routinely collected manufacturing data rather than relying solely on surface appearance.

More recently, the literature has increasingly shifted toward deep learning-based defect detection [11]. Recent studies report strong performance improvements through architectural optimization and lightweight backbones [12, 13]. For example, Wan *et al.* [14], Carvalho *et al.* [15], and Lu *et al.* [16] optimized YOLOv5 models and reported good performances. In addition, Chen *et al.* [17] proposed a multi-scale defect detection model for 3D-printed ceramic components. A recent contribution by Diao *et al.* [18] introduced a GhostNet-based lightweight YOLOv10s variant.

Across both machine learning and deep learning studies, class imbalance and partial overlap among defect categories are repeatedly identified as major obstacles in multi-class settings. Dominant classes (e.g., crack-related defects) tend to be easier to separate, whereas rarer or process-dependent defects are harder to distinguish. This makes class-sensitive evaluation essential, since aggregate accuracy alone can mask systematic errors on minority but operationally important defect categories. In this context, evaluation protocols that combine class-wise performance reporting with transparent decision rules become particularly important for industrial adoption.

Overall, the literature indicates that ceramic defect analysis has progressed from classical image processing to modern deep learning detectors and production-data–driven decision support. Nevertheless, challenges related to operational variability, class imbalance, and deployability remain. Building on these insights, the present study focuses on defect evaluation using a large-scale, real production dataset under industrial class imbalance.

3. Materials and methods

In this study, automated defect classification is treated as an evaluation process rather than a purely predictive task. Accordingly, the methodological design is guided by principles of measurement reliability, uncertainty management, and output stability.

In ceramic sanitaryware manufacturing, uncertainty arises from subjective visual judgments, overlapping defect mechanisms, and variability in production processes. These factors influence both the consistency of reference labels and the reliability of model predictions. To limit their effects, probabilistic calibration and ensemble aggregation are applied to stabilize output distributions and improve repeatability under different operating conditions. Within this context, classification confidence is viewed as a quantitative indicator of measurement reliability, which allows decision thresholds to be adjusted in line with operational risk considerations.

3.1 Data source

This study is based on real operational production data collected from a ceramic sanitaryware manufacturing company operating in Türkiye. The dataset comprises 11,071 production records obtained under routine industrial conditions and reflects authentic process variability. Each record corresponds to a single product unit and includes numerical and categorical variables that describe material properties, process parameters, and production conditions.

From a quality control standpoint, each product's defect condition is treated as the target quantity, whereas production and process variables provide indirect evidence for its assessment. Data were collected through routine monitoring practices, without artificial modification, resampling, or synthetic augmentation. This strategy helps maintain the authenticity of the production setting and supports the practical relevance of the analysis.

Defect labels were first assigned during routine quality control activities by experienced inspectors. When uncertainties arose, the samples were re-examined by a senior supervisor, and the final decision was reached through mutual agreement. This two-step annotation procedure enhances the consistency of the reference labels used for model training and testing. To further reinforce the reliability of the labeling process, a randomly selected subset of records was additionally reviewed by an independent domain expert in order to check for potential inconsistencies and borderline cases.

Although formal inter-annotator agreement metrics were not recorded, the two-stage inspection procedure, consensus-based validation of uncertain cases, and additional independent review of a subset of samples were designed to enhance label consistency under real production conditions. This approach reflects common industrial quality-control practice, where operational feasibility often limits the use of formal annotation protocols.

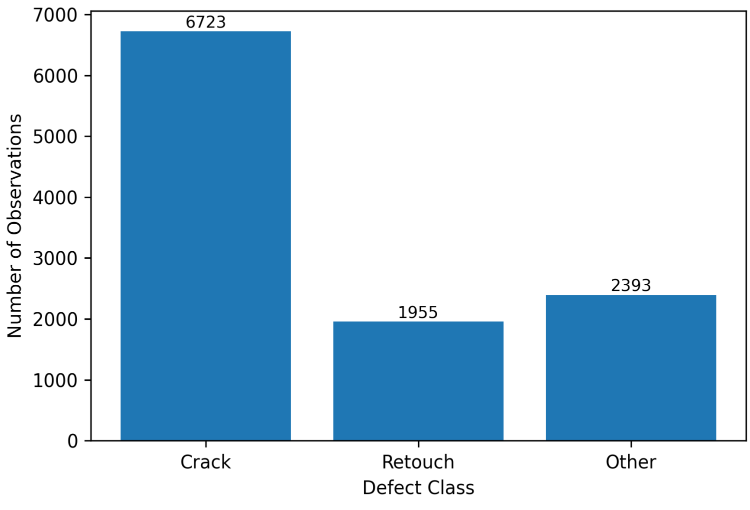

Fig. 1 Distribution of defect classes in the production dataset

Because several defect categories were represented by relatively few observations, the original labels were consolidated into three operationally meaningful classes: Crack, Retouch, and Other. The Other category includes air bubbles, pinholes, deformation defects, and assembly-related imperfections. Although these defect types differ in their physical characteristics, they were combined due to their comparable implications for downstream quality control, particularly with respect to rework and rejection decisions. In practice, these defects are mainly identified through the same visual inspection routines and production monitoring mechanisms, which further supports their joint treatment within a unified class structure. In addition, the dataset displays a pronounced class imbalance. As shown in Fig. 1, Crack defects form the majority class with 6,723 instances, while Retouch and Other are represented by 1,955 and 2,393 samples, respectively. This uneven distribution reflects realistic production conditions but poses considerable challenges for model training and performance assessment. Accordingly, class-aware evaluation metrics and probability calibration are emphasized throughout the analysis.

3.2 Feature preparation and candidate models

The dataset includes seven process- and material-related input variables representing indirect measurement channels of defect formation: five categorical variables (product category, kiln type, clay type, casting area, and drying chamber) and two continuous variables (thixotropy and mold age). Product category denotes the five main sanitaryware groups produced on the line (washbasin, toilet, bidet, urinal, and shower tray). Kiln type reflects four firing configurations used in production. Clay type distinguishes two material formulations with different raw-material composition and firing behavior (a fine-grained, high-density formulation for high-temperature firing and smooth-surface products, and a coarser, lower-density formulation typically used for thicker-walled products). The casting area refers to the specific location where casting and shaping are performed, covering eight distinct areas within the production line. The drying chamber denotes the controlled pre-firing drying environment, consisting of seven chambers, whose temperature and humidity conditions are known to affect crack formation and deformation. Furthermore, thixotropy is a rheological property that describes material flow behavior under mechanical stress. It is commonly expected to influence surface-related defects through its effects on application and adhesion. Mold age indicates the number of casting cycles a mold has completed and is linked to wear-related changes that may impact geometric stability and surface quality.

During data preprocessing, numerical features were standardized to have zero mean and unit variance in order to improve numerical stability and maintain comparability across learning algorithms. Categorical variables were encoded using one-hot representations for multinomial logistic regression and support vector machines, whereas CatBoost relied on its built-in permutation-based encoding scheme to process categorical information efficiently.

Three supervised learning models reflecting complementary modeling strategies were examined: multinomial logistic regression, support vector machines with a radial basis function kernel, and CatBoost. Logistic regression was adopted as a linear probabilistic baseline and trained with class-weighted loss and L2 regularization. Support vector machines employed a soft-margin formulation to capture nonlinear decision boundaries, with kernel and penalty parameters tuned via cross-validation. CatBoost was used to model complex nonlinear interactions among process variables, incorporating early stopping and regularization mechanisms to limit model complexity.

To exploit the complementary strengths of individual classifiers, a probability-based voting ensemble was constructed (see Fig. 2). Let $p_m^{(k)}$ denote the posterior probability assigned to class k by model m. The ensemble prediction is obtained as

$$\hat{y} = \arg\max_{k \in \kappa} \sum_{m=1}^{M} w_m p_m^{(k)}, \tag{1}$$

where w_m represents the weight associated with model m, and M denotes the number of base classifiers. Both equal-weight and performance-informed weighting strategies were examined. In the performance-informed scheme, weights were proportional to validation Macro-F1 scores and normalized to sum to unity.

To improve the reliability of posterior probability estimates, post-hoc probability calibration was applied using Platt scaling (sigmoid method) implemented via the CalibratedClassifierCV framework on validation data prior to ensemble aggregation. This calibration step ensures that probabilistic outputs are comparable across models and supports transparent and stable decision-making.

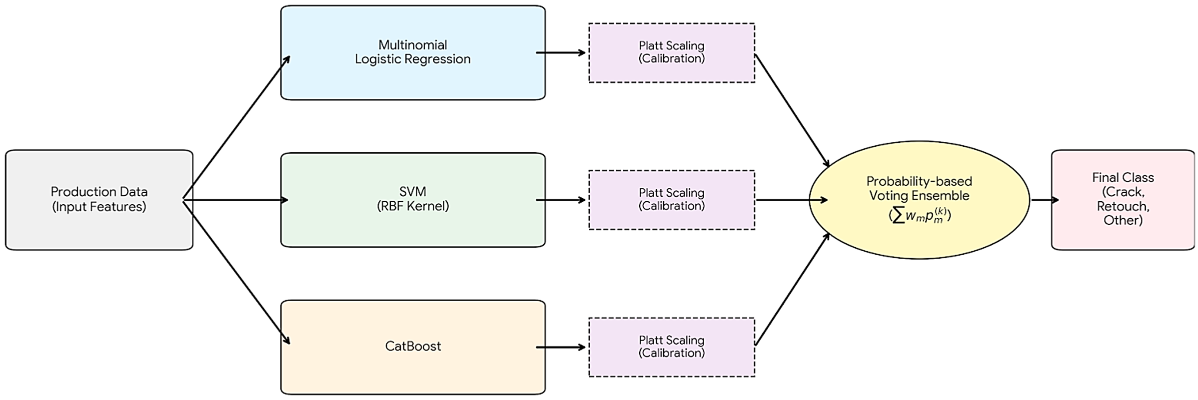

Fig. 2 Architecture of the proposed probability-based voting ensemble model

3.3 Experimental design and performance evaluation

The dataset was split into training and test sets using stratified sampling to preserve class proportions; 25 % of the observations were reserved as an external test set. No data from the test set was used during feature engineering, probability calibration, or hyperparameter tuning. Within the training set, hyperparameter optimization was performed using Optuna with 100 trials under a 5-fold cross-validation scheme. All experiments were conducted using a fixed random seed to ensure reproducibility. Macro-F1 was selected as the primary optimization objective to reflect balanced performance across imbalanced classes.

To support a more production-oriented interpretation of model outputs, probability-based decision behavior was examined through calibrated probability outputs and class-wise performance reporting. All preprocessing and modeling steps were implemented within a unified computational pipeline; stochastic components were controlled via fixed random seeds, and the complete experimental configuration was documented to support reproducibility. Calibration quality was evaluated on validation data to ensure that probabilistic outputs used in ensemble aggregation are comparable and decision-relevant.

4. Results and discussion

This section presents and discusses the empirical findings obtained from the comparative evaluation of individual classifiers and the proposed probability-based ensemble using real production data. The primary objective is to assess classification performance under industrial class imbalance and to examine the reliability of model outputs from a production-oriented perspective.

Table 1 reports the overall predictive performance of multinomial logistic regression, support vector machines, CatBoost, and the proposed ensemble on the external test set. As shown in Table 1, logistic regression and support vector machines achieve competitive Macro-F1 scores, while CatBoost exhibits slightly lower generalization performance in the present production-data context. The ensemble outperforms individual models in terms of balanced performance, indicating that probabilistic aggregation contributes to improved robustness under heterogeneous operating conditions. The proposed ensemble not only improves Macro-F1 performance but also reduces inter-fold performance variance and enhances probability calibration, leading to more stable and reliable decision support in production settings.

Table 1 Overall classification performance of individual models and the proposed ensemble

Model	Accuracy	Macro precision	Macro recall	Macro F1
Logistic regression	0.7695	0.7376	0.6946	0.7126
SVM (RBF)	0.7673	**0.7445**	0.6797	0.7047
CatBoost	0.7565	0.7260	0.6602	0.6838
Voting ensemble	**0.7731**	0.7345	**0.7099**	**0.7211**

Table 2 Class-wise precision, recall, and F1-score for each classifier on the test dataset

Class	Model	Precision	Recall	F1-Score
Crack	Logistic Regression	0.8011	0.8721	0.8351
	SVM (RBF)	0.7869	**0.8876**	0.8342
	CatBoost	0.7839	0.8870	0.8323
	Voting Ensemble	**0.8158**	0.8614	**0.8380**
Retouch	Logistic Regression	0.6939	0.5562	0.6175
	SVM (RBF)	**0.7155**	0.5194	0.6019
	CatBoost	0.7003	0.4683	0.5613
	Voting Ensemble	0.6608	**0.6094**	**0.6340**
Other	Logistic Regression	0.7179	0.6555	0.6853
	SVM (RBF)	**0.7311**	0.6321	0.6780
	CatBoost	0.6939	0.6254	0.6579
	Voting Ensemble	0.7249	**0.6589**	**0.6912**

Table 3 Confusion matrix of the proposed ensemble classifier on the test dataset

True / Predicted	Crack	Retouch	Other
Crack	1448	115	118
Retouch	161	298	30
Other	116	38	394

Detailed class-wise precision, recall, and F1-scores are summarized in Table 2. The results indicate that Crack defects are consistently detected with high accuracy across all models, reflecting their distinctive process-related signatures and higher frequency in the dataset. In contrast, Retouch defects exhibit substantially lower recall values and higher misclassification rates, primarily due to their partial overlap with both Crack and Other categories. This behavior highlights the intrinsic difficulty of reliably distinguishing retouch-related imperfections based solely on routinely collected production variables.

The confusion matrix corresponding to the ensemble classifier is presented in Table 3. Systematic confusion between the Retouch and Other classes is clearly observed. This pattern confirms that these categories share overlapping operational characteristics. From a quality-system perspective, such misclassification patterns represent a critical source of uncertainty, as they may lead to unstable rework decisions and inconsistent quality assessments in high-throughput production environments.

To support meaningful probability-based aggregation, posterior class probabilities were calibrated using Platt scaling prior to ensemble integration. The improved consistency of probabilistic outputs contributes to more stable decision behavior, reinforcing the interpretation of model predictions as measurement-informed indicators rather than purely statistical classifications. Moreover, after calibration, the Brier score decreased from 0.244 to 0.197, and the expected calibration error was reduced from 0.098 to 0.049.

Overall, the results demonstrate that defect evaluation based on routinely collected production data can be formulated as a reliable quality control support process when appropriate preprocessing, probability calibration, and ensemble integration are employed. Rather than maximizing aggregate accuracy alone, the proposed evaluation strategy emphasizes balanced performance, decision stability, and operational interpretability. These characteristics are essential for deploying machine learning-assisted inspection systems in industrial quality control, where consistent and transparent measurement outcomes are required for effective process control and continuous improvement.

In addition, the relevance of individual features was further examined using permutation-based importance analysis to assess their contribution to decision reliability. Since the proposed ensemble integrates posterior probabilities from multiple classifiers, direct attribution of decision behavior to individual input variables is not straightforward. Therefore, a model-agnostic permutation approach was adopted to quantify the marginal contribution of each feature to classification performance in terms of changes in Macro-F1 score.

The analysis indicates that thixotropy represents the most influential process variable, as its random perturbation resulted in the largest performance degradation. This finding highlights the critical role of material flowability and shaping behavior in defect formation, which is consistent with established ceramic manufacturing knowledge. Clay type and drying chamber variables likewise showed high importance scores. This indicates that defect occurrence is closely associated with material properties and environmental process conditions throughout production.

Mold age and product category exhibited moderate importance, indicating that these variables may influence defect formation indirectly through their effects on process stability and geometric properties. By contrast, casting location showed limited individual impact, and kiln-related variables displayed negligible importance, pointing to relatively uniform firing conditions within the examined production system. Namely, these results suggest that the proposed production-oriented framework not only enhances classification reliability but also yields practical insights into the key process pathways associated with defect generation.

Beyond overall performance, the results highlight that defect evaluation based on routinely collected production variables is inherently constrained by operational variability and partially unobserved factors. In particular, systematic confusion between Retouch and Other suggests that the available process descriptors may not fully capture the latent mechanisms driving subtle surface-related imperfections. This limitation is consistent with the measurement interpretation of the problem, where repeatability and decision reliability depend not only on the learning algorithm but also on the representativeness of observation channels and the stability of the operating context. The proposed calibration-aware ensemble reduces bias toward dominant classes and provides more stable probability-based decisions; however, residual errors indicate that additional contextual variables (e.g., operator-, shift-, and maintenance-related metadata) may be required to further improve reliability in high-throughput settings. The stability of feature rankings was further verified across cross-validation folds, indicating that the identified key process variables consistently contribute to classification reliability.

From an operational perspective, the observed improvement in class-balanced performance and minority-class recall contributes to more consistent rework decisions, reduces the risk of systematic misclassification, and supports more stable quality control processes in high-throughput production environments.

Furthermore, it should be noted that the dataset used in this study originates from a single production facility. While this limits the direct generalizability of model-specific results, the proposed framework is designed as a transferable decision-support approach. In different industrial settings, the same methodology can be applied by re-calibrating model parameters and adapting feature representations to local production conditions. In addition, the aggregation of multiple defect types into the Other category, while necessary to ensure statistical robustness, may lead to a loss of defect-specific information. However, this grouping reflects practical industrial decision-making, where defects with similar operational consequences are often treated collectively. Future work will focus on more fine-grained defect categorization as larger and more balanced datasets become available.

Lastly, the temporal stability of the model across different production periods, work shifts, and potential process drift could not be directly assessed within the scope of the current dataset. This aspect is critical for long-term deployment of data-driven quality control systems and will be addressed in future studies through time-aware validation and continuous monitoring approaches.

5. Conclusion

This study investigated production-data-based multi-class defect evaluation in ceramic sanitaryware manufacturing from a practical industrial quality control perspective. Using 11,071 real-world industrial observations and a two-stage quality-control labeling procedure, multinomial logistic regression, SVM (RBF), CatBoost, and a calibrated probability-based voting ensemble were evaluated under severe class imbalance. The results show that probability calibration and ensemble aggregation improve balanced performance and decision consistency, supporting the use of routinely collected process data as an indirect data-driven input for defect assessment.

Model-agnostic permutation analysis further indicates that defect behavior is primarily driven by material and process-condition variables, with thixotropy, clay type, and drying environment emerging as key contributors to classification reliability. At the same time, remaining misclassifications—especially between structurally similar defect categories—suggest that part of the decision uncertainty arises from latent and context-dependent factors that are not explicitly represented in the current dataset. Future work will focus on integrating operator-, shift-, and maintenance-related metadata to explicitly model contextual uncertainty and enhance decision consistency and reliability in high-throughput environments.

Acknowledgement

The authors would like to thank the industrial partner for providing access to real production data and supporting the development of this study. The authors also acknowledge the valuable insights obtained from practical manufacturing experience during the design of the experimental framework.

References

[1] Ozcetin, E., Ozturk, G., Ozturk, Z.K., Kasimbeyli, R., Kasimbeyli, N. (2023). A decision support system for consolidated distribution of a ceramic sanitary ware company, *Expert Systems with Applications*, Vol. 213, Part A, Article No. 118785, doi: 10.1016/j.eswa.2022.118785.

[2] Forrest, A.K. (2003). Surface defect detection on ceramics, In: Graves, M., Batchelor, B. (eds.), *Machine vision for the inspection of natural products*, Springer, London, United Kingdom, 191-213, doi: 10.1007/1-85233-853-9_6.

[3] Elbehiery, H., Hefnawy, A., Elewa, M. (2005). Surface defects detection for ceramic tiles using image processing and morphological techniques, In: *Proceedings of the Third World Enformatika Conference,* Istanbul, Turkey, 158-162.

[4] Hocenski, Z., Keser, T., Baumgartner, A. (2007). A simple and efficient method for ceramic tile surface defects detection, In: *Proceedings of 2007 IEEE International Symposium on Industrial Electronics*, Vigo, Spain, 1606-1611, doi: 10.1109/ISIE.2007.4374844.

[5] Karimi, M.H., Asemani, D. (2014). Surface defect detection in tiling industries using digital image processing methods: Analysis and evaluation, *ISA Transactions*, Vol. 53, No. 3, 834-844, doi: 10.1016/j.isatra.2013.11.015.

[6] Dengiz, O., Smith, A.E., Nettleship, I. (2006). Two-stage data mining for flaw identification in ceramics manufacture, *International Journal of Production Research*, Vol. 44, No. 14, 2839-2851, doi: 10.1080/00207540500534454.

[7] Kesharaju, M., Nagarajah, R. (2017). Particle swarm optimization approach to defect detection in armour ceramics, *Ultrasonics*, Vol. 75, 124-131, doi: 10.1016/j.ultras.2016.07.008.

[8] Hanzaei, S.H., Afshar, A., Barazandeh, F. (2017). Automatic detection and classification of the ceramic tiles' surface defects, *Pattern Recognition*, Vol. 66, 174-189, doi: 10.1016/j.patcog.2016.11.021.

[9] Rahaman, G.M.A., Hossain, M.M. (2009). Automatic defect detection and classification technique from image: A special case using ceramic tiles, *arXiv,* doi: 10.48550/arXiv.0906.3770.

[10] Kumru, M. (2013). Assessing the visual quality of sanitary ware by fuzzy logic, *Applied Soft Computing*, Vol. 13, No. 8, 3646-3656, doi: 10.1016/j.asoc.2013.03.012.

[11] Zhou, M., Wu, T., Xia, Z., He, B., Kong, L.B., Su, H. (2025). Research progress in deep learning for ceramics surface defect detection, *Measurement*, Vol. 242, Part B, Article No. 115956, doi: 10.1016/j.measurement.2024.115956.

[12] Abbas, S., Moumen, H., Abbas, F. (2023). Efficient method using attention based convolutional neural networks for ceramic tiles defect classification, *Revue d'Intelligence Artificielle*, Vol. 37, No. 1, 53-62, doi: 10.18280/ria.370108.

[13] Fei, Z., Zhao, R., Xin, K., Chen, G., Xiang, X. (2026). Defect detection of ceramic substrates based on improved YOLOv9 algorithm, *International Journal of Machine Learning and Cybernetics*, Vol. 17, No. 1, Article No. 29, doi: 10.1007/s13042-025-02835-2.

[14] Wan, G., Fang, H., Wang, D., Yan, J., Xie, B. (2022). Ceramic tile surface defect detection based on deep learning, *Ceramics International*, Vol. 48, No. 8, 11085-11093, doi: 10.1016/j.ceramint.2021.12.328.

[15] Carvalho, R., Morgado, A.C., Gonçalves, J., Kumar, A., Rolo, A.G.E.S., Carreira, R., Soares, F. (2023). Computer-aided visual inspection of glass-coated tableware ceramics for multi-class defect detection, *Applied Sciences*, Vol. 13, No. 21, Article No. 11708, doi: 10.3390/app132111708.

[16] Lu, Q., Lin, J., Luo, L., Zhang, Y., Zhu, W. (2022). A supervised approach for automated surface defect detection in ceramic tile quality control, *Advanced Engineering Informatics*, Vol. 53, Article No. 101692, doi: 10.1016/j.aei.2022.101692.

[17] Chen, W., Zou, B., Yang, G., Zheng, Q., Lei, T., Huang, C., Liu, J., Li, L. (2024). A real-time detection system for multiscale surface defects of 3D printed ceramic parts based on deep learning, *Ceramics International*, Vol. 50, No. 8, 13101-13112, doi: 10.1016/j.ceramint.2024.01.220.

[18] Diao, J., Wei, H., Zhou, Y., Diao, Z. (2025). A deep learning-based algorithm for ceramic product defect detection, *Applied Sciences*, Vol. 15, No. 12, Article No. 6641, doi: 10.3390/app15126641.