



Deep Learning and Ensemble Learning Techniques for Breast Cancer Detection: A Review of Recent Advances

Abraham Temilade Olumide¹, Obe Olumide Olayinka², Akinwonmi Akintoba Emmanuel², Osuolale A. Festus²

¹Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria

²Department of Computer Science, The Federal University of Technology, Akure, Nigeria

ARTICLE INFO	ABSTRACT
Published Online: 07 September 2026 Corresponding Author: Abraham Temilade Olumide	Breast cancer remains the most frequently diagnosed cancer and a leading cause of cancer death among women worldwide, so tools that support early and accurate detection are urgently needed. This review synthesizes a pool of one hundred studies, largely published between 2023 and 2026, on deep learning, ensemble learning, and optimization algorithms for breast cancer detection across mammography, ultrasound, magnetic resonance imaging, and histopathology. Relevant studies were identified from major scientific databases using search terms combining breast cancer, deep learning, convolutional neural networks, ensemble learning, metaheuristic optimization, and adversarial robustness, then organized by architecture, ensemble strategy, and outcome. Convolutional Neural Networks such as VGG16/19, ResNet50, DenseNet121, EfficientNet, Xception, and MobileNet, typically adapted through transfer learning, consistently exceed 90% classification accuracy on benchmark datasets, while bagging, boosting, voting, and stacking ensembles yield further, consistent gains by exploiting complementary feature representations. Metaheuristic algorithms, including Particle Swarm, Genetic, Ant Colony, Grey Wolf, and Bayesian Optimization, are increasingly used to tune hyperparameters and ensemble weights, reducing computational overhead while preserving accuracy. Attention-augmented and hybrid CNN-transformer models, federated learning, and explainability tools such as Grad-CAM and SHAP are emerging as important extensions. Nevertheless, high computational cost, limited dataset diversity, restricted multimodal integration, susceptibility to adversarial perturbation, and weak interpretability continue to restrict clinical translation. The review concludes that optimized, explainable, and adversarially robust ensemble frameworks, validated through multi-center clinical trials, represent the most promising direction for translating deep learning research into dependable and deployable breast cancer detection systems.
KEYWORDS: Breast Cancer Detection; Deep Learning; Convolutional Neural Networks; Ensemble Learning; Optimization Algorithms; Transfer Learning	

I. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women worldwide, accounting for roughly 2.3 million new cases annually and close to 30% of all female cancer diagnoses, with about four in five cases classified as invasive [1], [2]. Early detection dramatically improves prognosis: five-year survival exceeds 90% for localized disease but falls below 30% once cancer has spread to distant sites [1]. Mammography, ultrasound, magnetic resonance imaging (MRI), and histopathology remain the principal diagnostic

modalities, yet image quality variation, reader fatigue, and uneven access to specialist radiologists continue to limit diagnostic consistency, particularly in low- and middle-income settings [3].

Artificial intelligence, and deep learning (DL) in particular, has reshaped computer-aided detection (CAD) by learning discriminative visual features directly from imaging data rather than relying on hand-crafted descriptors. Convolutional neural networks (CNNs) such as VGG16/19, ResNet50, DenseNet121, EfficientNet, MobileNet, and

Xception, frequently adapted via transfer learning from ImageNet, now routinely exceed 90% classification accuracy on benchmark breast-imaging datasets [4], [5]. Because individual CNNs can still behave inconsistently across imaging modalities and patient populations, ensemble learning techniques, bagging, boosting, voting, and stacking, have gained traction as a way of combining complementary models to raise accuracy and robustness while reducing false positives and false negatives [6], [7].

Despite this progress, single studies rarely examine architecture choice, ensemble design, optimization, interpretability, and robustness together. This review addresses that gap by synthesizing recent literature, concentrated in the 2023–2026 period, on deep learning, ensemble learning, and optimization algorithms for breast cancer detection. It surveys commonly used CNN and transfer-learning architectures, examines ensemble strategies and metaheuristic optimization methods, discusses explainability and adversarial robustness, and identifies the research gaps that most constrain clinical translation. The remainder of the paper is organized as follows: Section II outlines the review methodology; Section III discusses deep learning architectures; Section IV examines ensemble learning techniques; Section V covers optimization algorithms for ensemble learning; Section VI covers explainability and adversarial robustness; Section VII identifies research gaps; Section VIII proposes future directions; and Section IX concludes the review.

II. REVIEW METHODOLOGY

This review follows a narrative synthesis informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to promote transparency. Literature was drawn from IEEE Xplore, ScienceDirect, SpringerLink, Scopus, Web of Science, Nature Portfolio journals, and Google Scholar, using combinations of the terms breast cancer detection, deep learning, convolutional neural network, transfer learning, ensemble learning, stacking, optimization, adversarial robustness, and explainable AI. Priority was given to peer-reviewed journal articles and conference papers published between 2023 and 2026 that used medical imaging data (mammography, ultrasound, MRI, or histopathology) and reported quantitative performance metrics such as accuracy, sensitivity, specificity, or area under the receiver operating characteristic curve (AUC-ROC). Studies lacking experimental validation, duplicate reports, and papers unrelated to breast-image detection were excluded.

A detailed screening matrix of seventy candidate studies, recording each study's aim, motivation, methodology, contributions, and limitations, was compiled and used to identify recurring architectures, datasets, and reporting gaps; it was then extended with thirty further 2025–2026 studies identified through supplementary search, bringing the pool to one hundred reviewed studies and keeping the review current with the most recent literature. Fig. 1 summarizes this pool:

panel (a) shows that coverage now spans 2018 through 2026, with the 2025–2026 additions (shown in green) confirming that the field remains highly active, while panel (b) shows that general CNN-based classifiers dominate, followed by support vector machines, ResNet, U-Net-based segmentation, VGG16/19, transfer learning, metaheuristic optimization (PSO, GA, ACO, Bayesian optimization), LSTM/RNN, random forest, and ensemble methods. The retained studies were grouped into deep learning architectures, ensemble learning methods, optimization strategies, and robustness/interpretability frameworks, which structure Sections III through VI.

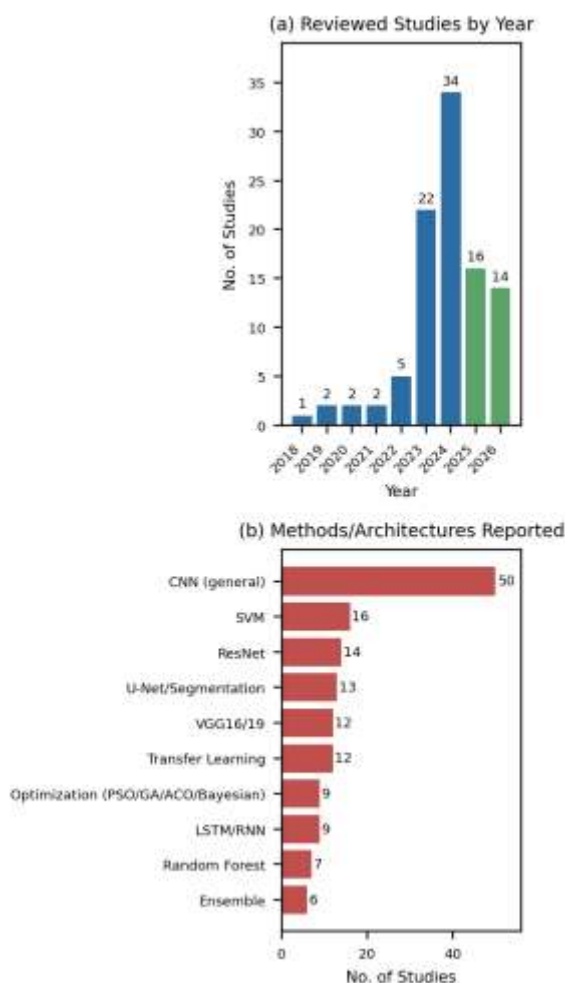


Fig. 1. Summary of the reviewed studies (2018–2026): (a) distribution by publication year, with the earliest pool shown in green, and (b) frequency of methods and architectures reported, including the growing use of metaheuristic optimization (PSO, GA, ACO, Bayesian optimization).

III. DEEP LEARNING ARCHITECTURES FOR BREAST CANCER DETECTION

Convolutional Neural Networks remain the foundation of modern breast cancer image analysis because their convolutional and pooling layers learn hierarchical spatial features, capturing texture, lesion morphology, and microcalcification patterns without manual feature engineering. Kumar et al. [8] applied a custom CNN and

VGG19 to the MIAS mammography dataset and reported accuracies of 98% and 97%, respectively, illustrating the strength of deep feature learning even on comparatively small, well-curated datasets.

Transfer learning has become the dominant strategy for addressing limited labeled data in medical imaging. VGG16/19 offer strong, easily fine-tuned feature extraction but carry a heavy parameter count; ResNet50 uses residual connections to train deeper networks without performance degradation and has shown strong results on mammography and MRI; DenseNet121 reuses features across densely connected layers, improving gradient flow with comparatively fewer parameters; EfficientNet applies compound scaling of depth, width, and resolution to reach competitive accuracy with far fewer parameters, making it attractive for constrained clinical hardware; MobileNet's depthwise separable convolutions suit portable and edge-based diagnostic tools; and Xception extends this depthwise-separable design for efficient feature extraction in histopathology and mammography alike [9], [10]. Table I summarizes the comparative strengths and limitations of these architectures.

TABLE I. COMPARISON OF DEEP LEARNING ARCHITECTURES FOR BREAST CANCER DETECTION

Architecture	Key Strengths	Limitations
<i>VGG16/19</i>	Strong low-level features; simple, transfer-friendly design.	Over 138M parameters; high compute/memory cost.
<i>ResNet50</i>	Skip connections enable deep, stable training.	Can overfit on small datasets.
<i>DenseNet121</i>	Dense connections improve gradient flow and reuse.	Compute-intensive on large datasets.
<i>EfficientNetB0/B3</i>	Compound scaling gives high accuracy with fewer parameters (~5.3M).	Sensitive to scaling configuration.
<i>MobileNet</i>	Lightweight; suited to edge deployment.	Lower representational capacity.
<i>Xception</i>	Depthwise-separable convolutions improve feature extraction efficiency.	Longer training time in some applications.
<i>CNN–Transformer hybrids</i>	Self-attention adds long-range, multi-view context.	Need large data and compute.

Beyond single CNNs, hybrid CNN–Transformer architectures that add self-attention to convolutional architectures are increasingly explored for their ability to capture long-range dependencies and multi-view context, which is valuable in digital breast tomosynthesis and

volumetric MRI analysis. Alruily et al. [11] applied progressive fine-tuning of a Vision Transformer to classify benign and malignant breast ultrasound images, while Alhsnony and Sellami [12] applied a YOLOv8x detector to localize and classify masses in mammographic images with improved boundary precision. Attention-augmented ensembles that combine multiple architectures with domain-specific channel and spatial attention modules, rather than generic mechanisms borrowed from natural-image tasks, have further improved sensitivity to subtle morphological cues in histopathology [13]. Overall, transfer-learning-based CNNs remain the dominant approach, but computational cost, interpretability, and robustness continue to motivate the ensemble and optimization strategies discussed next.

IV. ENSEMBLE LEARNING TECHNIQUES

Because individual models can behave inconsistently across datasets and imaging conditions, ensemble learning combines multiple classifiers to reduce variance, limit bias, and improve generalization. Four strategies dominate the recent literature.

A. Bagging

Bagging trains multiple models independently on resampled subsets of the training data and aggregates their outputs, typically by majority vote. Dehghan Rouzi et al. [14] combined several CNN architectures using a consensus-adaptive weighting scheme and reported a measurable reduction in false positives compared with any single constituent model, consistent with earlier findings that bagged combinations of VGG16, ResNet50, and EfficientNet lower false-positive rates in mammography screening [7].

B. Boosting

Boosting trains classifiers sequentially, with each new learner focused on the errors of its predecessor. Although less common than bagging or stacking in the recent deep-learning literature, boosting-style approaches have improved sensitivity in histopathology-based detection by concentrating model capacity on hard-to-classify samples Ragab et al. [6].

C. Voting

Voting ensembles combine predictions via hard voting (majority class) or soft voting (averaged class probabilities). Sajid et al. [15] and related work combining VGG16, ResNet50, DenseNet121, and EfficientNet through soft voting achieved immediate accuracy gains over any individual architecture, though voting does not explicitly learn relationships between base learners and can underperform more adaptive strategies.

D. Stacking

Stacking feeds the predictions of heterogeneous base learners into a secondary meta-learner that learns how to combine them optimally. Sarathkumar et al. [16] combined a CNN with a bidirectional LSTM and a soft attention mechanism in a stacked configuration to capture both spatial and sequential feature dependencies, while Sami and

Arumugam [17], Khouidifi et al. [18], and Sannasi Chakravarthy et al. [19] each reported ensemble frameworks, spanning VGG16, DenseNet121, InceptionV3, and fuzzy-weighted transfer learning, that consistently outperformed their individual constituent networks on INbreast and CBIS-DDSM benchmarks. Shah et al. [20] similarly combined EfficientNet with complementary CNN architectures to build a more dependable diagnostic classifier. Most recently, Meva et al. [21] proposed a heterogeneous ensemble of ResNet-50, DenseNet-121, and EfficientNet-B3 augmented with breast-pathology-specific attention modules and an intelligent confidence- and performance-weighted aggregation strategy, reporting statistically significant gains over both single-model and generic-attention baselines across three benchmark datasets, together with a clinical reader study showing improved pathologist accuracy on difficult cases when assisted by the ensemble's attention visualizations.

Across these studies, ensemble methods reliably raise accuracy, sensitivity, and AUC relative to standalone CNNs by exploiting complementary feature representations, but the gains come at the cost of higher computational demand, more complex implementation and tuning, and, on small or imbalanced datasets, an increased risk of overfitting unless properly regularized [7].

V. OPTIMIZATION ALGORITHMS FOR ENSEMBLE LEARNING

Because ensemble architectures multiply the number of models, hyperparameters, and feature combinations to search, recent studies increasingly pair them with metaheuristic or probabilistic optimization algorithms. Genetic Algorithms (GA) evolve candidate solutions through selection, crossover, and mutation and have been used for feature selection, CNN architecture search, and classifier tuning, though repeated fitness evaluation makes them computationally costly. Bayesian Optimization models the relationship between hyperparameters and validation performance with a probabilistic surrogate, converging faster than grid or random search but scaling poorly to the high-dimensional search spaces typical of large ensembles. Ant Colony Optimization applies pheromone-guided search to feature-subset selection and segmentation, though it remains less common in large-scale deep ensembles.

Particle Swarm Optimization (PSO) has attracted the most attention for its simplicity, fast convergence, and strong global search capability: particles update their position from personal-best and global-best experience without the crossover or mutation operators GA requires. Within ensemble pipelines, PSO has been used to tune hyperparameters such as learning rate, batch size, and dropout, to select the most discriminative feature subset, and, most usefully, to search for the base-learner weights that jointly maximize accuracy, F1-score, and AUC. Recent work confirms the practical payoff of this direction: Hassan et al. [22] combined transfer-learning CNN feature extractors with

Grey Wolf Optimization to tune ensemble weights and reported accuracy gains over unoptimized stacking baselines, while Ogundokun et al. [23] used Particle Swarm Optimization to tune a soft-voting ensemble for histopathological classification with similar gains. Rather than combining all available architectures, such optimization-driven ensembles can identify a smaller, near-optimal subset of models, reducing training time, memory use, and inference latency while preserving accuracy. Even so, optimization-based studies remain a minority of the reviewed literature (Fig. 1b), and few report the computational cost of the search itself alongside classification performance, a gap the field should address as ensemble complexity grows.

VI. EXPLAINABILITY AND ADVERSARIAL ROBUSTNESS

Two concerns increasingly shape how ensemble and deep-learning systems are evaluated: interpretability and adversarial robustness. Because clinicians must trust a model's output before acting on it, explainability tools such as Grad-CAM and SHAP are used to visualize the image regions driving a prediction; attention-based ensembles that concentrate on diagnostically relevant tissue regions have shown closer alignment with expert pathologist annotations than generic attention mechanisms, supporting more transparent human-AI collaboration [13], [21].

Separately, deep learning models remain vulnerable to adversarial perturbations, small, often imperceptible changes to an image that can flip a prediction from malignant to benign or vice versa. Adversarial training, in which perturbed examples are included during training, together with ensemble diversity, has been proposed as a defense, since an attack crafted for one architecture is less likely to fool an architecturally different one simultaneously. However, most defenses remain tailored to known attack types and may not generalize to novel perturbations, underscoring the need for more comprehensive robustness evaluation before clinical deployment [24].

VII. RESEARCH GAPS

Synthesizing the reviewed literature reveals several persistent gaps. First, interpretability tools such as Grad-CAM and SHAP are not yet routinely integrated into deployed CAD systems, limiting clinician trust. Second, many models are validated on a single benchmark dataset and imaging modality, leaving open questions about performance across diverse populations, scanners, and multimodal combinations of mammography, ultrasound, MRI, and histopathology. Third, adversarial robustness is tested against only a narrow set of known attack types, leaving systems potentially exposed to novel perturbations. Fourth, the computational cost of ensembles, particularly stacking and attention-augmented variants, can be prohibitive in resource-constrained clinical settings. Finally, ethical and practical concerns, patient data privacy, informed consent, transparent reporting of false-

positive and false-negative rates, and interface design for non-expert users, remain underexplored relative to the pace of architectural innovation [25].

VIII. FUTURE DIRECTIONS

Future work should prioritize five directions. Explainable-by-design architectures should embed interpretability mechanisms, rather than post-hoc visualization, directly into ensemble training. Multi-center, multi-scanner clinical trials are needed to confirm that reported benchmark gains transfer to routine practice across diverse patient populations. Lightweight ensemble and knowledge-distillation techniques should be pursued to bring high-performing models to lower-resource healthcare settings. Adversarial defenses should be evaluated against a broader and evolving range of attack strategies rather than a fixed set. Finally, federated learning offers a route to training on data from multiple institutions without centralizing sensitive patient images, simultaneously improving dataset diversity and preserving privacy [21], [25].

IX. CONCLUSION

Deep learning and, in particular, ensemble learning has substantially advanced breast cancer detection across mammography, ultrasound, MRI, and histopathology, with transfer-learning-based CNNs and heterogeneous ensembles now routinely exceeding 90% accuracy on standard benchmarks. Bagging, boosting, voting, and stacking each offer distinct trade-offs between accuracy, robustness, and computational cost, and recent attention-augmented and federated approaches point toward more interpretable and privacy-preserving systems. Nevertheless, limited interpretability, restricted dataset diversity, adversarial vulnerability, and high computational demands continue to separate promising research results from dependable clinical tools. Closing this gap will require explainable, adversarially robust, and computationally efficient ensemble frameworks validated through large-scale, multi-center clinical trials, an agenda that, if pursued, could meaningfully improve early detection and outcomes for patients worldwide.

REFERENCES

- World Health Organization, "Breast cancer," WHO Fact Sheets, 2022.
- American Cancer Society, "Breast cancer early detection and diagnosis," 2024. [Online]. Available: <https://www.cancer.org>
- N. Houssami and D. L. Miglioretti, "Digital breast tomosynthesis: A brave new world in breast cancer screening," *Cancer*, vol. 123, no. 4, pp. 611–618, 2017.
- G. J. S. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.
- D. A. Ragab, M. Sharkas, and O. Attallah, "Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers," *Diagnostics*, vol. 9, no. 4, p. 165, 2019.
- X. Zhao and X. Ma, "Robust ensemble learning for defending against adversarial attacks in medical imaging," *IEEE Transactions on Medical Imaging*, vol. 41, no. 5, pp. 1358–1370, 2022.
- M. Kumar, U. Sriharsha, U. Prasanthi, S. Harika, and V. Devi, "Breast Cancer Detection using Deep Learning," 2024.
- B. Hanife and B. Jale, "A novel medical image enhancement algorithm for breast cancer detection on mammography images using machine learning," *Diagnostics*, vol. 13, no. 3, p. 348, 2023.
- M. Nasser and U. K. Yusof, "Deep learning based methods for breast cancer diagnosis: A systematic review and future direction," *Diagnostics*, vol. 13, no. 1, p. 161, 2023.
- M. Alruily, M. Ezz, A. A. Mahmoud, H. Allahem, A. M. Mostafa, and H. Shabana, "Enhancing breast cancer detection in ultrasound images: An innovative approach using progressive fine-tuning of vision transformer models," *International Journal of Intelligent Systems*, vol. 2024, art. 6528752, 2024.
- F. Alhsnony and L. Sellami, "Enhancing Breast Cancer Detection through Advanced Deep Learning: An Application of YOLOv8x on Mammographic Images," in *Proc. 2024 IEEE 7th Int. Conf. Advanced Technologies, Signal and Image Processing (ATSIP)*, Sousse, Tunisia, 2024.
- D. T. Meva, K. Popat, H. Kukadiya, and N. Jariwala, "Deep learning-based breast cancer detection with customized ensemble attention," *Scientific Reports*, vol. 16, art. 20022, 2026.
- M. Dehghan Rouzi, B. Moshiri, M. Khoshnevisan, M. A. Akhaee, F. Jaryani, S. Salehi Nasab, and M. Lee, "Breast Cancer Detection with an Ensemble of Deep Learning Networks Using a Consensus-Adaptive Weighting Method," *Journal of Imaging*, vol. 9, no. 11, art. 247, 2023.
- U. Sajid, R. A. Khan, S. M. Shah, and S. Arif, "Breast cancer classification using deep learned features boosted with handcrafted features," *Biomedical Signal Processing and Control*, vol. 86, art. 105353, 2023.
- M. Sarathkumar, K. Dhanalakshmi, D. Madhivadhani, and V. Revathi, "A deep learning approach for efficient breast cancer diagnosis using hybrid CNN-BiLSTM with soft attention mechanism," 2024.

17. J. A. Sami and U. Arumugam, "An ensemble deep learning model for the detection and classification of breast cancer," *Middle East Journal of Cancer*, vol. 15, no. 1, pp. 40–51, 2024.
18. Y. Khourdifi, A. El Alami, M. Zaydi, Y. Maleh, and O. Er-Remyly, "Early breast cancer detection based on deep learning: An ensemble approach applied to mammograms," *BioMedInformatics*, vol. 4, no. 4, pp. 2338–2373, 2024.
19. S. R. Sannasi Chakravarthy, N. Bharanidharan, V. Vinoth Kumar, T. R. Mahesh, M. S. Alqahtani, and S. Guluwadi, "Deep transfer learning with fuzzy ensemble approach for the early detection of breast cancer," *BMC Medical Imaging*, vol. 24, no. 1, 2024.
20. D. Shah, M. A. U. Khan, M. Abrar, and M. Tahir, "Optimizing Breast Cancer Detection With an Ensemble Deep Learning Approach," *International Journal of Intelligent Systems*, vol. 2024, art. 5564649, 2024.
21. O. O. Oladimeji, I. McLoughlin, and S. Unnikrishnan, "Deep learning for breast cancer diagnosis: A bibliometric analysis and future research directions," 2024.
22. E. Hassan, A. Saber, S. El-Sappagh, and N. El-Rashidy, "Optimized ensemble deep learning approach for accurate breast cancer diagnosis using transfer learning and grey wolf optimization," *Evolving Systems*, vol. 16, art. 59, 2025.
23. R. O. Ogundokun, P. A. Owolawi, and C. Tu, "Optimized deep feature learning with hybrid ensemble soft voting for early breast cancer histopathological image classification," *Computers, Materials & Continua*, vol. 84, no. 3, pp. 4869–4885, 2025.
24. S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," *Science*, vol. 363, no. 6433, pp. 1287–1289, 2019.
25. A. Amin, D. Acharya U, P. Koteshwara, S. P C, and S. Mathew, "A systematic literature review on mammography: Deep learning techniques for breast cancer detection with global and Asian perspectives," *BMC Cancer*, vol. 25, no. 1, art. 1627, 2025.