International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

INTERNATIONAL JOURNAL OF INNOVATION AND INDUSTRIAL REVOLUTION (IJIREV)
www.ijirev.com

ENHANCING BREAST CANCER PREDICTION WITH AUTOML TPOT CONFIGURATIONS FOR MAMMOGRAPHIC BREAST CANCER CLASSIFICATION

Faikah Awang@ Ismail[1,2], Muhammad Khalis Abdul Karim[2*], Mohd Mustafa Awang Kechik2, Izdihar Kamal[2], Zahurin Ismail[3], Nurul Syazatul Filzah Mohamad Suchah[2]

[1] School of Biology, Faculty of Applied Sciences, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Kuala Pilah, 72000, Malaysia
Email: faikah7450@uitm.edu.my
[2] Department of Physics, Faculty of Sciences, Universiti Putra Malaysia, 43400, Malaysia
Email: mkhalis@upm.edu.my, mmak@upm.edu.my, izdiharkamal@upm.edu.my, syazafilzah32@gmail.com
[3] Department of Radiology, Hospital Putrajaya, Ministry of Health Malaysia, 62250, Putrajaya
Email: drzahurin@hpj.gov.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 27.03.2025
Revised date: 14.04.2025
Accepted date: 15.05.2025
Published date: 05.06.2025

To cite this document:

Ismail, F. A., Karim, M. K. A., Kechik, M. M. A., Kamal, I., Ismail, Z., Suchah, M. N. S. F (2025). Enhancing Breast Cancer Prediction with AutoML TPOT Configurations For Mammographic Breast Cancer Classification. *International Journal of Innovation and Industrial Revolution, 7* (21), 01-15.

DOI: 10.35631/IJIREV.721001

Abstract:

Breast cancer remains a leading cause of death among women worldwide. Early detection using mammographic imaging improves patients' outcomes; however, its reliability is heavily dependent on radiologist's expertise, often leading to variability and misdiagnosis. This study explores the potential of Automated Machine Learning (AutoML) on enhancing breast cancer prediction by comparing three configurations Tree-Based Pipeline Optimization Tool (TPOT): Default, Light and Sparse. 244 mammography images were obtained from two database access through The Cancer Imaging Archive (TCIA). Image pre-processing was using MATLAB R2022a, USA, employing Contrast Limited Adaptive Histogram Equalization (CLAHE) for image enhancement and Active Contour Method (ACM) for image segmentation of region of interest (ROI). subsequently enabling the extraction of radiomic features. These extracted features were then used to train and test three TPOT configurations; TPOT Defaults, TPOT Light and TPOT Sparse via Python version 3.9. The classification models later were evaluated for accuracy, sensitivity and specificity to ascertain the model's efficiency in distinguishing between benign and malignant breast cancer. 37 radiomic features: six First-Order Statistical features, 21Gray-Level Co-occurrence Matrix (GLCM) Texture features and ten Shape-Based Features were extracted from sample images. The TPOT Default configuration achieved the highest accuracy of 0.735 (CI95%: 0.611-0.859), with a sensitivity of 0.760 (CI95%:

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

0.642- 0.878), and precision of 0.731 (CI95%:0.607-0.855) outperforming both TPOT Light, accuracy: 0.633 (CI95%:0.498 -0.768), sensitivity: 0.667 (CI95%: 0.537- 0.797), precision; 0.615 (CI95%:0.485-0.745) and TPOT Sparse with accuracy: 0.673 (CI95%: 0.543-0.803), sensitivity, 0.653 (CI95%: 0.521-0.785) and precision; 0.708 (CI95%:0.587 -0.829). These results demonstrate that the TPOT Default configuration delivers the most reliable classification, highlighting AutoML's potential as a clinical decision support tool. By reducing manual feature engineering and improving diagnostic accuracy, AutoML could significantly streamline breast cancer detection and improve outcomes in radiological practice.

Keywords:

Breast Cancer, Automated Machine Learning, Machine Learning, Mammography, Radiomic, TPOT

Introduction

Breast cancer remains as one of the leading causes of cancer-related deaths among women worldwide. Despite advancements in medical imaging and diagnostic approaches, challenges persist in the early detection of breast cancer. Mammography, considered the gold standard imaging modality, plays a critical role in breast cancer screening and diagnosis (Duffy et al., 2021). However, interpreting mammographic images can be challenging due to variability in tumour appearance, breast tissue density and the subtle characteristic of early-stage tumors. Even experienced radiologist may encounter difficulties, often resulting in false positives, missed diagnoses or unnecessary biopsies (Huynh et al., 2023).

The incorporation of artificial intelligence (AI) and machine learning (ML) into medical imaging analysis has introduced promising advancement. Radiomics, which involves extracting quantitative data from medical images, when combined with ML, has significantly enhanced diagnostic accuracy. However, conventional ML approaches typically require extensive technical expertise, posing a challenge for healthcare providers who lack in data science (Dhillon et al., 2022). To address this limitation, automated machine learning (AutoML) frameworks, such as the TPOT have emerged. TPOT automates key processes including model selection, hyperparameter tuning and feature engineering though genetic programming, enabling non-experts to develop predictive models with minimal coding experience (Moore et al., 2023). This study aims to evaluate the built-in TPOT configurations, namely Defaults, Light and Sparse in classifying mammographic breast tumours as benign or malignant using the radiomic features.

Literature Review

Breast Cancer and Tumor Classification

Breast cancer is the most diagnosed cancer in women worldwide and remains a leading cause of cancer-related mortality particularly in industrialized nations. This is largely attributed to lifestyle factors such as poor diet, smoking, stress and sedentary behavior. Hormonal influences, including late-age pregnancies and hormone replacement therapy intake, further contribute to the rising burden of this cancer (Smolarz et al., 2022).Socioeconomic factors also play a critical role in breast cancer epidemiology, influencing access to screening, early detection and timely treatment. In many developing countries, limited healthcare infrastructure often results in late-stage diagnosis and poorer outcomes, thereby elevating mortality rates.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

Given these trends, effective prevention strategies, early screening programs and broader healthcare access are essential to mitigate the global burden of breast cancer (Iacoviello et al., 2021). The prognosis of breast cancer is heavily influenced by the type of tumor whether benign or malignant as each requires different treatment approaches and has distinct survival outcomes. Benign tumors are non-cancerous growth that only remain localized and do not metastasis to other parts of the body. These tumors require regular surveillance and personalized care plans. Malignant tumors, in contrast, are aggressive and capable of metastasizing. Their classification is determined by histological features, hormone receptor status and genetic markers which in turn inform treatment strategies and (Vrigazova, 2020) . Early detection significantly improves clinical outcomes, and advancements in imaging technologies such as digital mammography, Magnetic Resonance Imaging (MRI) and ultrasound have enhanced tumor classification accuracy (Kalaiyarasi et al., 2020) .Nonetheless, malignant tumors continue to pose a serious health threat, underscoring the urgent need for more advanced, accurate and accessible diagnostic tools.

Imaging and Image Interpretation

Digital mammography is a widely used imaging modality for breast cancer detection and screening. Unlike traditional film-based methods, it utilizes electronic detectors to convert X-rays into digital images, providing clearer visualization and facilitating easier analysis. This digital format enables image enhancements such as magnification and contrast adjustment, which assist radiologists in identifying abnormalities more effectively (Zhou et al., 2022). When integrated with computer-aided detection (CAD) system and deep learning algorithms, digital mammography can detect subtle indicators of cancer, including small, non-palpable lesions such as ductal carcinoma in situ (DCIS). These technological advancements have contributed to increased detection rates, reduced recall rates and a decrease in unnecessary biopsies (Farber et al., 2021) . However, despite this improvement, overdiagnosis remains a concern as image interpretation largely depends on radiologist's experience, training and skill in recognizing subtle abnormalities in breast tissue (Lee et al., 2021). To promote consistency and accuracy, structured reporting systems like BI-RADS are employed to standardize tumors classification. Furthermore, advancement in AI and radiomics have shown potential in supporting radiologists by automating lesion detection and categorization. These tools enhance diagnostic accuracy, particularly in reducing false-positive (FP) and false-negative (FN) findings, both of which significantly influence patient management. FP results may cause unnecessary to anxiety and lead to unnecessary procedures, while FN can delay treatment and worsening the prognosis (Schaffter et al., 2020). Studies have demonstrated that by combining AI with radiologist interpretation can improve diagnosis specificity, thereby reducing both FP and FN rates. Ongoing research and training in image interpretation, along with continued integration of AI integration, are essential for improving the precision of breast cancer screening and minimizing diagnostic errors (Brunetti et al., 2025).

Automated Machine Learning TPOT

AutoML particularly TPOT, is revolutionizing breast cancer diagnosis. TPOT automates model selections, hyperparameter tunings and ML pipelines construction through genetic programming. This approach eliminates the need for extensive manual effort and technical expertise typically required in conventional ML workflows. By discovering complex patterns and feature interactions that may be overlooked by human analysts, TPOT enhance predictive performance and minimizes bias during model development. In breast cancer imaging, TPOT has demonstrated strong capabilities in distinguishing between benign and malignant tumors,

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

even within complex, high-dimensional datasets such as mammographic images (Radzi et al., 2021). TPOT has been shown to outperform conventional ML models in terms of accuracy, sensitivity and specificity, making it a valuable tool for clinicians (Akaramuthalvi & Palaniswamy, 2021; Yuan et al., 2024). Its application can significantly reduce FP and FN rates, enabling healthcare providers to make faster and more accurate diagnostic decisions. By simplifying the ML pipeline and enhancing accessibility to users without specialized programming knowledge, AutoML framework like TPOT have the potential to support better decision-making and ultimately improve patient outcomes (Le et al., 2020; Yuan et al., 2024).

Methodology

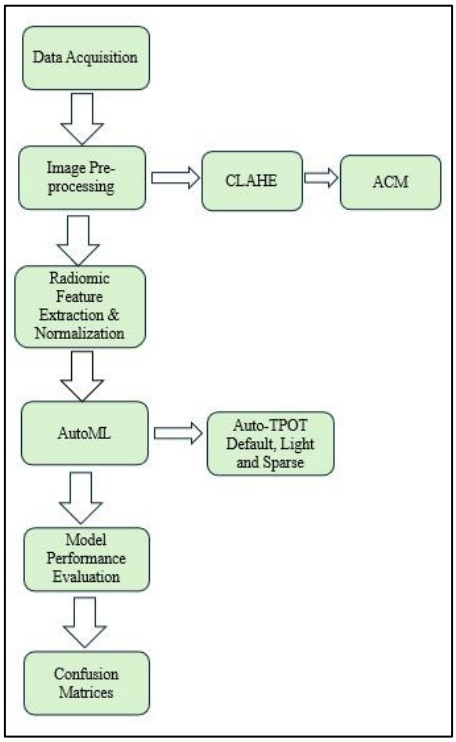

Figure 1: Workflow Used for Study

Dataset Acquisition

The complete workflow is illustrated in Figure 1. This retrospective study utilized the mammographic image sourced from two publicly available databased; the Curated Breast Imaging Subset of Digital Database for Screening Mammography and Digital Database for Screening Mammography (CBIS-DDSM), both accessible via The Cancer Imaging Archive (TCIA), https://www.cancerimagingarchive.net/collection/cbis-ddsm/. A total of 244 mammogram images were selected, comprising both Craniocaudal (CC) and Mediolateral Oblique (MLO) views. These included 125 benign tumours and 119 malignant tumors, all pathologically confirmed by the radiologist. Prior to analysis, all images were converted into DICOM format for standardization. Image enhancement was performed using CLAHE, followed by ACM for accurate segmentation of the region of intereset (ROI). Both processes were executed in MATLAB R2022a software. After image segmentation, 37 radiomic features; six first-order, 21 GLCM texture second order and ten shape-based were extracted. These features served as inputs for model development using the AutoML TPOT framework, which

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001
was implemented in Python version 3.9. The TPOT configuration were then used for automated model construction, hyperparameter tuning and evaluation of performance metrics.

Image Pre-processing

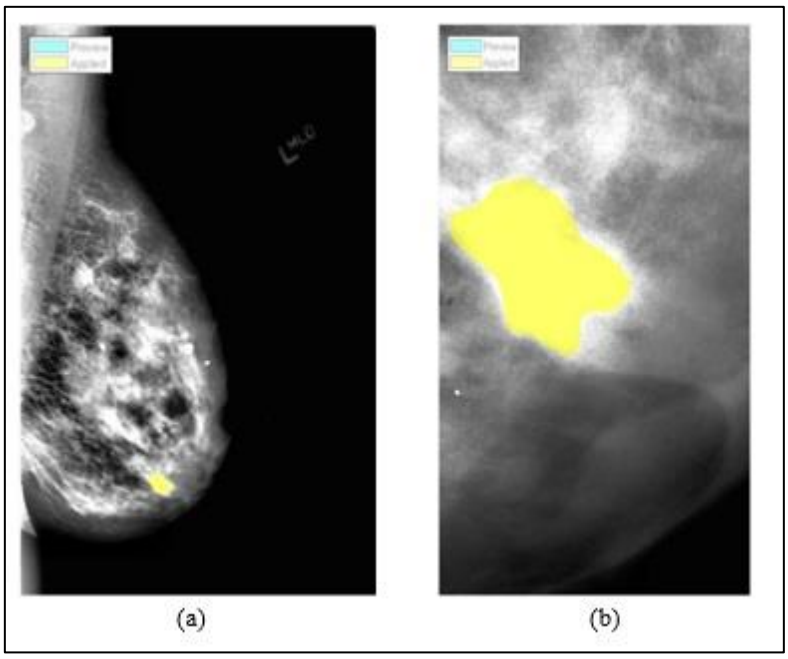

Figure 2: (a) Image Enhancement via CLAHE and (b) Image Segmentation via ACM

Figure 2 illustrates the application CLAHE and ACM in the preprocessing of mammographic images. CLAHE was implemented using MATLAB R2022a, where images were first loaded into the software environment for image enhancement. CLAHE enhances local contrast while suppressing noise amplification, thereby improving the visibility of structural details such as dense breast tissue and microcalcifications. This technique is particularly effective for highlighting subtle variations in tissue density, thus enhancing the accuracy of subsequent image segmentation and radiomic feature extraction (Dheeba et al., 2014). CLAHE operates by dividing the image into localized regions, or "tiles," and applying histogram equalization independently to each tile. In this study, the clip limit was manually set to 0.02, and the tile grid size was left at the default setting of 8 × 8, which provided optimal contrast enhancement. Following this, the ACM command was employed to segment the region of interest (ROI) that contained the breast mass. ACM iteratively refines the contour of the segmented area by minimizing an energy function, allowing for precise delineation of tumor boundaries, especially those with irregular shapes (Alshamrani et al., 2023). The number of iterations was set to 150 to ensure adequate convergence. This method proved effective in distinguishing tumors from surrounding breast tissue and capturing complex mass contours (Iqbal et al., 2020). To validate segmentation accuracy, the results were reviewed and confirmed by both a mammographer and a radiologist, ensuring that all relevant lesions were accurately segmented prior to radiomic feature extraction.

Radiomic Feature Extraction and Normalization

Following the segmentation process, radiomic features were extracted from the regions of interest (ROIs) to quantify differences between benign and malignant breast tumors. Three primary categories of radiomic features were analyzed: first-order statistical features, GLCM

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

textural features, and shape-based features. First-order statistical features characterize the overall intensity distribution within the segmented lesion and include metrics such as mean, variance, skewness, kurtosis, entropy, and energy (Tagliafico et al., 2020). GLCM features evaluate the spatial relationships between pixel intensities, producing texture-based metrics like contrast, correlation, energy, and homogeneity, which are useful in identifying tissue heterogeneity (Srivastava et al., 2020). Shape-based features such as area, perimeter, eccentricity, convex area, and solidity were computed to assess the morphological characteristics of the breast masses. To ensure the extracted features were consistent across different imaging systems, Min-Max Scaling was applied to normalize the data. This method scales all feature values to a standardized range of [0, 1), thereby preventing features with larger magnitudes from disproportionately influencing the model. Standardization is particularly important in medical imaging datasets where images may originate from various mammography machines with differing output characteristics. Min-Max Scaling enhances machine learning performance by reducing biases introduced by feature scale disparities and improving algorithm convergence speed (Izonin et al., 2022). Additionally, the relative contribution of each radiomic feature to the classification outcome was examined to identify the most influential predictors for differentiating between benign and malignant lesions.

Application of TPOT for Breast Cancer Classification

The classification of breast masses was conducted using the AutoML TPOT framework, which automates the selection of machine learning models, feature processing methods, and hyperparameter tuning. TPOT utilizes genetic programming to optimize classification pipelines, enabling the identification of high-performing models without the need for manual configuration or expert intervention (Olson et al., 2016). In this study, three TPOT configurations were evaluated; TPOT Default, TPOT Light and TPOT Sparse. TPOT Default performs comprehensive search across various preprocessing strategies, feature selection techniques, and classification algorithms. While this makes it the most robust option, it is also computationally intensive (Radzi et al., 2021). TPOT Light, on the other hand, is optimized for efficiency. It reduces the number of models and feature selection methods used, making it well-suited for environments with limited computational resources. Lastly, TPOT Sparse is designed specifically for datasets that may include missing values. It incorporates feature selection strategies that are better equipped to handle incomplete data, ensuring more reliable performance in such scenarios (Olson et al., 2016).

Model Performance Evaluation

Each TPOT configuration was trained and tested using an 80:20 data split, and its classification performance was evaluated using several metrics, including accuracy, sensitivity, precision, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) analysis. Accuracy measures the overall proportion of correctly classified cases, providing a general indication of model performance. Sensitivity (or recall) evaluates the model's ability to correctly identify malignant tumors, which is critical in minimizing false-negative results and ensuring early detection (Radzi et al., 2021). In contrast, precision measures the model's capacity to correctly classify benign tumors, thereby reducing unnecessary biopsies and false-positive outcomes. The ROC curve offers a visual representation of the trade-off between sensitivity and precision across different classification thresholds. The AUC (Area Under the Curve) provides a single numerical value summarizing this performance, where a value closer to 1.0 indicates a higher discriminatory power of the model in distinguishing between benign and malignant breast

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001
tumors (Fusco et al., 2021). Collectively, these metrics provide a comprehensive view of each model's diagnostic capability and clinical applicability.

Classification of Lesions via Confusion Matrices

Confusion matrices were generated for each TPOT configuration to visualize classification performance and highlight instances of misclassification. These matrices provided a clearer understanding of each model's diagnostic accuracy, offering insights into the specific types of errors, False Positive and False Negative, made during classification. The results of this analysis underscore the potential of AutoML-based breast cancer classification, demonstrating how TPOT effectively optimizes machine learning pipelines for clinical application. By automating feature selection and model optimization, TPOT reduces dependence on manual, trial-and-error approaches, thereby improving both the efficiency and accuracy of breast cancer detection. Using an 80:20 training-to-testing ratio, a total of 49 lesions from the full dataset of 244 samples, comprising a mix of benign and malignant cases that were randomly selected for the confusion matrix evaluation. In this context, True Positives (TP) refer to correctly identified malignant tumors, while True Negatives (TN) represent correctly identified benign cases. Conversely, False Positives (FP) are benign lesions misclassified as malignant, potentially leading to unnecessary biopsies or interventions, and False Negatives (FN) are malignant tumors that were incorrectly identified as benign, which may delay critical treatment. These matrix components provide a detailed breakdown of model performance, aiding in a comprehensive assessment of clinical reliability.

Results

The findings indicate that CLAHE significantly enhanced the contrast of mammographic images by improving the visibility of mass structures while minimizing noise amplification. Compared to the original grayscale images, those processed with CLAHE exhibited enhanced differentiation between normal and abnormal tissue regions, which is critical for accurate mass detection (Radzi et al., 2020). This enhancement was particularly beneficial for visualizing microcalcifications and low-contrast tumors that may be overlooked during manual mammographic interpretation (Dheeba et al., 2014). Following contrast enhancement, ACM segmentation was applied to accurately isolate ROIs containing breast lesions. The results demonstrated that ACM effectively delineated tumor margins, especially for lesions with irregular shapes. This method, which operates through iterative energy minimization, adapted well to the variability in tumor morphology, thereby improving segmentation accuracy. Accuracy was verified through a combination of manual ROI selection and ACM refinement, ensuring alignment with the natural contours of the lesions. The segmented ROIs showed a high level of agreement with radiologist annotations, confirming the robustness of the ACM approach (Iqbal et al., 2020). Visual comparisons revealed that ACM reliably captured the edges of both benign and malignant tumors, including cases with partially obscured boundaries. Malignant tumors, often characterized by spiculated and irregular borders, were clearly separated from surrounding tissue, enabling more precise feature extraction. Similarly, benign tumors, which typically present smoother, rounded shapes, were also effectively segmented. The accuracy of segmentation was critical for radiomic analysis, as errors in delineation could distort the values of texture and shape-based features used for classification.

A total of 37 radiomic features were extracted from the mammographic images, comprising six first-order statistical features, 21 GLCM texture features, and ten shape-based features. The analysis revealed that malignant tumors exhibited higher entropy and skewness, reflecting their

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

increased heterogeneity and irregular tissue structures. These variations indicate that malignant tumors tend to be more disorganized and structurally complex than benign counterparts. The GLCM texture features effectively captured the textural differences between tumor types. For instance, contrast, which measures the degree of intensity variation, was significantly higher in malignant tumors due to their irregular internal structures. In contrast, homogeneity, which assesses the uniformity of pixel intensity, was more prominent in benign tumors, which typically display smoother and more consistent textures. Regarding shape-based features, malignant tumors frequently demonstrated irregular, non-circular shapes and spiculated borders which is a key characteristics associated with invasive growth. Benign tumors, on the other hand, were generally rounder and more symmetric. These distinct radiomic patterns highlight the potential of feature-based image analysis in accurately differentiating between benign and malignant breast lesions.

The model performance evaluation results highlight the different classification efficacy among three TPOT configurations used in this study, namely, TPOT Default, TPOT Light and TPOT Sparse. Table 1 shows the performance metric for all TPOT configurations in differentiating benign and malignant breast cancer in sample images.

Table 1: Model Performance for Three TPOTS Configuration

TPOT Configuration	Accuracy	CI95%	Sensitivity	CI95%	Precision	CI95%
TPOT Default	0.735	(0.611,0.859)	0.760	(0.642,0.878)	0.731	(0.607,0.855)
TPOT Light	0.633	(0.498,0.768)	0.667	(0.537,0.797)	0.615	(0.485,0.745)
TPOT Sparse	0.673	(0.543,0.803)	0.653	(0.521,0.785)	0.708	(0.587,0.829)

TPOT Default consistently produced the most reliable results across all evaluated metrics. It achieved the highest accuracy of 0.7325 (CI95%: 0.611–0.859) and sensitivity of 0.760 (CI95%: 0.642–0.878), demonstrating its strong capability to accurately distinguish between benign and malignant breast lesions. In contrast, TPOT Light recorded the lowest performance, with an accuracy of 0.633 (CI95%: 0.498–0.768) and sensitivity of 0.667 (CI95%: 0.537–0.797), suggesting that its streamlined configuration may compromise diagnostic reliability. TPOT Sparse showed slightly better performance than TPOT Light, achieving an accuracy of 0.673 (CI95%: 0.543–0.803) and sensitivity of 0.653 (CI95%: 0.521–0.785). However, it still fell short of matching TPOT Default in overall classification effectiveness. These findings affirm that TPOT Default provides the most robust and accurate performance, making it the most suitable configuration for early and reliable detection of breast malignancies.

In terms of precision, TPOT Default again demonstrated superior performance, achieving a value of 0.731 (CI95%: 0.607–0.855), reflecting its strong ability to accurately predict malignant cases without excessive false positives. TPOT Sparse followed with a precision of 0.708 (CI95%: 0.587–0.829), while TPOT Light recorded the lowest precision at 0.615 (CI95%: 0.486–0.745). Although TPOT Sparse showed respectable precision, it did not outperform TPOT Default. Moreover, its reduced sensitivity raises concerns about the risk of overlooking actual cancer cases, which could limit its clinical utility. These findings suggest that TPOT Default not only provides the highest precision but also achieves a crucial balance with sensitivity, reinforcing its status as the most reliable and diagnostically robust configuration for breast cancer classification using mammographic images in this study.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

Based on Figure 3, the ROC-AUC curves reveal notable differences in the performance of each TPOT configuration in distinguishing between benign and malignant breast cancer cases. The TPOT Default configuration attained the highest AUC value of 0.732, indicating strong discriminatory power in identifying positive (malignant) versus negative (benign) outcomes. The ROC curve for TPOT Default exhibits a steep initial rise, which reflects a higher true positive rate (TPR) at lower false positive rates (FPR), a critical trait in medical diagnostics, where minimizing false negatives is vital to avoid missed cancer diagnoses. This performance underscores TPOT Default's effectiveness in supporting early and accurate detection in clinical breast cancer screening.

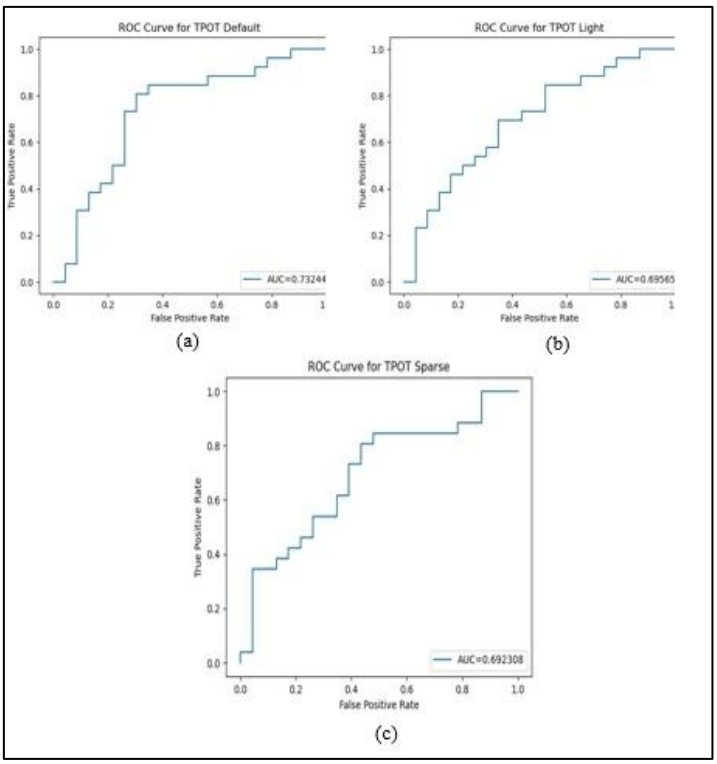

Figure 3: ROC-AUC for (a) TPOT Default with AUC 0.732, (b) TPOT Light with AUC 0.696 and (c) TPOT Sparse with AUC 0.692

In contrast, the TPOT Light and TPOT Sparse configurations recorded lower AUC values of 0.695 and 0.692, respectively. Their ROC curves display a more gradual slope, indicating a reduced ability to discriminate between benign and malignant cases, particularly at moderate false positive rates. Although the performance differences are not extreme, these results highlight the consistency of TPOT Default in delivering more reliable and stable classification boundaries. Among the three configurations, TPOT Default offers the most balanced and accurate classification, as evidenced by its superior AUC score. This further reinforces earlier findings, positioning TPOT Default as the most suitable model for clinical applications where diagnostic precision and early detection are critical.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

Table 2: Confusion Matrix Summary for All TPOT Configuration for Tumor Classification

TPOT Configuration	TP	TN	FP	FN	Total Correct	Total Misclassified
TPOT Default	19	17	6	7	36	13
TPOT Light	16	15	8	10	31	18
TPOT Sparse	21	12	11	5	33	16

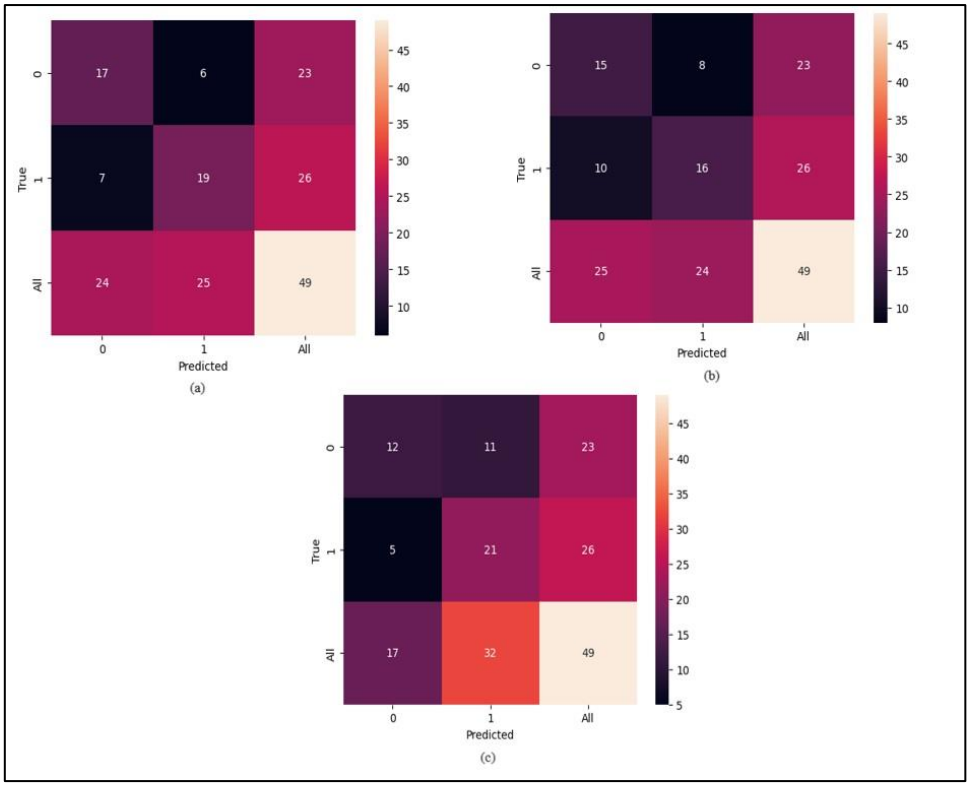

Figure 4: Confusion Matrices for (a) TPOT Default, (b) TPOT Light and (c) TPOT Sparse

Table 2 and Figure 4 present a comprehensive breakdown of the classification performance for each TPOT configuration. TPOT Default achieved the highest accuracy, correctly classifying 36 out of 49 test samples, which included 19 true positives (TP) and 17 true negatives (TN), along with 6 false positives (FP) and 7 false negatives (FN). TPOT Sparse followed with 33 correct classifications, recording the highest number of true positives (21). However, it also produced the highest number of false positives (11), indicating a tendency to overpredict malignancy. In contrast, TPOT Light exhibited the lowest overall performance, correctly classifying only 31 cases while misclassifying 18. It also registered the fewest true positives (16) and the second-lowest number of true negatives (15). These findings underscore the superior balance and accuracy of TPOT Default, while highlighting the limitations of TPOT Light and the over-sensitivity of TPOT Sparse.

These results further support the earlier performance metrics, reinforcing that TPOT Default demonstrates the most balanced classification capability, effectively identifying both malignant and benign tumors while maintaining low false positive (FP) and false negative (FN) rates. In

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

contrast, TPOT Sparse, although more aggressive in detecting malignancies, as evidenced by its highest true positive (TP) count which also produced the highest FP rate, which may lead to unnecessary follow-ups and clinical interventions. TPOT Light showed the weakest diagnostic reliability across all performance indicators, highlighting its limited effectiveness in this application.

To evaluate whether the differences in classification performance among the models were statistically significant, pairwise McNemar's tests were conducted using the outcomes from the confusion matrices. These tests assessed whether the classification discrepancies between model pairs were due to chance or indicative of meaningful differences. The results of the pairwise comparisons are summarized in Table 3.

Table 3: Pairwise McNemar's Test Results Comparing Classification Performance Between TPOT Configurations

Comparison	McNemar's X^2	p-value	Significant (α=0.05)
TPOT Defaults vs TPOT Light	3.20	0.074	No
TPOT Defaults vs TPOT Sparse	5.14	0.023	Yes
TPOT Light vs TPOT Sparse	0.125	0.724	No

The comparison between TPOT Default and TPOT Light yielded a McNemar's chi-square value of 3.20 with a p-value of 0.074, indicating that although TPOT Default showed superior accuracy and sensitivity, the difference between the two configurations was not statistically significant. In contrast, the comparison between TPOT Default and TPOT Sparse resulted in a statistically significant difference, with a chi-square value of 5.14 and a p-value of 0.023. This confirms that TPOT Default significantly outperformed TPOT Sparse in classification reliability, particularly in accurately identifying both malignant and benign cases.

The final comparison, between TPOT Light and TPOT Sparse, produced a chi-square value of 0.125 and a p-value of 0.724, suggesting no meaningful difference in classification performance between these two configurations. While TPOT Light offered slightly improved efficiency over Sparse, neither matched the classification performance of the Default configuration. Collectively, these findings support the conclusion that TPOT Default is the most statistically robust and clinically viable AutoML model for breast cancer classification using mammographic images in this study.

Discussions

Early detection of breast cancer is critical, as it significantly improves patient survival rates and facilitates more effective treatment planning. Mammographic imaging is considered the gold standard for breast cancer screening due to its ability to visualize soft tissue structures within the breast. However, its diagnostic accuracy can be limited by low image contrast, particularly when detecting small masses or microcalcifications in dense breast tissue, which may lead to missed or incorrect diagnoses. Manual interpretation of mammograms by radiologists is also subject to interobserver variability and contributes to the growing workload of medical professionals.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

To address these challenges, this study implemented a comprehensive workflow that includes image preprocessing techniques such as contrast enhancement and segmentation, along with the application of Automated Machine Learning (AutoML) for breast cancer classification. These methods aim to improve diagnostic accuracy, reduce subjectivity, and enhance overall efficiency in the detection process. By integrating image enhancement and machine-driven analysis, the workflow supports more consistent and reliable decision-making in breast cancer diagnosis.

In this study, the effectiveness of AutoML in predicting breast cancer was evaluated by analyzing the performance of various built-in configurations of the TPOT on enhanced and segmented mammographic images for distinguishing between benign and malignant tumors. Image enhancement was shown to significantly contribute to improved classification accuracy. Prior research has demonstrated that histogram-based enhancement techniques, such as CLAHE and Histogram Intensity Windowing (HIW), enhance mass detection, particularly when combined with AI models(Alshamrani et al., 2023; Nguyen-Tat et al., 2025; Radzi et al., 2021). Furthermore, Active Contour Method (ACM) segmentation was crucial for the precise extraction of relevant features, playing a vital role in minimizing tumor misclassification. Accurate segmentation directly supports the optimization of machine learning model performance in breast cancer diagnostics by ensuring that the extracted radiomic features are representative of the actual lesion (Mahmood et al., 2024). The integration of CLAHE, ACM, and AutoML within a unified workflow demonstrates strong potential for enhancing diagnostic accuracy and supporting clinical decision-making.

The findings from radiomic feature extraction in this study confirmed that malignant tumors exhibit distinct characteristics, such as higher heterogeneity and spiculated borders, whereas benign tumors typically display more homogeneous textures and smoother contours. These observations are consistent with previous research, reinforcing the significance of radiomic features in breast cancer detection (Radzi et al., 2020). For instance, prior research found that radiomic-based analysis of digital mammography could effectively distinguish between benign and malignant breast tumors, achieving an AUC of 0.934 (CI95%: 0.898–0.971) in the training set and 0.901 (CI95%: 0.835–0.961) in the test set (Wang et al., 2022). These AUC values are notably higher than those observed in the current study, which may be attributed to differences in dataset size and diversity. The larger dataset used by Wang et al., 2022 likely contributed to the enhanced generalizability and robustness of their model. Nevertheless, the current study provides valuable insights into the potential of radiomic features combined with AutoML for effective breast cancer classification in mammographic imaging.

When compared with recent advancements in AutoML, studies have demonstrated that TPOT consistently outperforms conventional machine learning models by automating feature selection and hyperparameter tuning (Rashed et al., 2023). Similarly, the integration of ensemble learning methods with AutoML has been shown to significantly enhance prediction accuracy in medical imaging applications (Imrie et al., 2025). These findings underscore the value of TPOT in streamlining the breast cancer classification process while minimizing human error. Despite the superior performance of TPOT Default, it does come with a trade-off which is higher computational cost and longer processing time, as it explores a wide range of generation paths and pipeline combinations. This observation is consistent with findings from previous research (Radzi et al., 2021). A notable limitation of this study is the class imbalance within the dataset. Among the 244 mammographic images analysed, 125 were benign (51.2%)

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

and 119 were malignant (48.8%), representing a relatively balanced distribution. However, even minor imbalances can skew the performance of machine learning models toward the majority class, especially when working with small datasets. This may impact the sensitivity of the model when applied to larger and more diverse populations. Moreover, the dataset was sourced from only two public databases, and although standardized preprocessing and feature extraction procedures were applied, the lack of external validation raises concerns about the generalizability of the findings. All models were trained and evaluated using data processed under controlled conditions from a limited range of imaging systems. As a result, it is uncertain how these AutoML model, particularly TPOT configurations would perform across diverse demographic groups, different clinical environments, or images acquired using varying mammography equipment and protocols. Therefore, future research should include external validation on multi-institutional datasets and testing with a variety of imaging devices to ensure the robustness and clinical applicability of AutoML findings in real-world settings.

Conclusion

Mammography remains the gold standard for early breast cancer detection. Integrating image preprocessing with AutoML, particularly TPOT pipelines, significantly enhances diagnostic accuracy and efficiency. Among the configurations, TPOT Default proved most reliable, achieving the highest accuracy, sensitivity, precision, and AUC. These findings highlight the potential of AutoML and radiomic analysis to support more precise and automated decision-making in clinical practice.

Acknowledgement

We would like to thank School of Biology, UiTM N. Sembilan and Department of Physics, Faculty of Sciences, UPM for their contribution in this research.

References

Akaramuthalvi, J. B., & Palaniswamy, S. (2021). Comparison of conventional and automated machine learning approaches for breast cancer prediction. *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, 1533–1537. https://doi.org/10.1109/ICIRCA51532.2021.9544863

Alshamrani, K., Alshamrani, H. A., Alqahtani, F. F., & Almutairi, B. S. (2023). Enhancement of mammographic images using histogram-based techniques for their classification using CNN. *Sensors*, *23*(1), 1–22. https://doi.org/10.3390/s23010235

Brunetti, N., Campi, C., Piana, M., Picone, I., Vercelli, C., Starovatskyi, O., Rescinito, G., Tosto, S., Garlaschi, A., Calabrese, M., & Tagliafico, A. S. (2025). A radiomic and clinical data-based risk model for malignancy prediction of breast BI-RADS 4A microcalcifications. *Clinical Breast Cancer*, 1–9. https://doi.org/10.1016/j.clbc.2025.01.006

Dheeba, J., Albert Singh, N., & Tamil Selvi, S. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics*, *49*, 45–52. https://doi.org/10.1016/j.jbi.2014.01.010

Dhillon, S. K., Ganggayah, M. D., Sinnadurai, S., Lio, P., & Taib, N. A. (2022). Theory and practice of integrating Machine Learning and conventional statistics in medical data analysis. In *Diagnostics* (Vol. 12, Issue 10, pp. 1–25). https://doi.org/10.3390/diagnostics12102526

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

Duffy, S. W., Tabár, L., Yen, A. M., Dean, P. B., Smith, R. A., Jonsson, H., Törnberg, S., Chiu, S. Y., Chen, S. L., Jen, H.-H., Ku, M. M., Hsu, C.-Y., Ahlgren, J., Maroni, R., Holmberg, L., & Chen, C. (2021). Beneficial effect of consecutive screening mammography examinations on mortality from breast cancer: A prospective study. *Radiology*, *299*(3), 541–547. https://doi.org/10.1148/radiol.2021203935

Farber, R., Houssami, N., Wortley, S., Jacklyn, G., Marinovich, M. L., McGeechan, K., Barratt, A., & Bell, K. (2021). Impact of full-field digital mammography versus film-screen mammography in population screening: A meta-analysis. *Journal of the National Cancer Institute*, *113*(1), 16–26. https://doi.org/10.1093/JNCI/DJAA080

Fusco, R., Granata, V., Raso, M. M., Vallone, P., De Rosa, A. P., Siani, C., Di Bonito, M., Petrillo, A., & Sansone, M. (2021). Blood oxygenation level dependent magnetic resonance imaging (MRI), dynamic contrast enhanced MRI and diffusion weighted MRI for benign and malignant breast cancer discrimination: A preliminary experience. *Cancers*, *13*(10), 1–15. https://doi.org/10.3390/cancers13102421

Huynh, H. N., Tran, A. T., & Tran, T. N. (2023). Region-of-Interest optimization for Deep-Learning-Based breast cancer detection in mammograms. *Applied Sciences*, *13*(12), 1–19. https://doi.org/10.3390/app13126894

Iacoviello, L., Bonaccio, M., de Gaetano, G., & Donati, M. B. (2021). Epidemiology of breast cancer, a paradigm of the "common soil" hypothesis. *Seminars in Cancer Biology*, *72*(December 2019), 4–10. https://doi.org/10.1016/j.semcancer.2020.02.010

Imrie, F., Denner, S., Brunschwig, L. S., Maier-hein, K., & Schaar, M. Van Der. (2025). Automated ensemble multimodal machine learning for healthcare. *IEEE Journal of Biomedical and Health Informatics*, 1–14. https://doi.org/10.1109/JBHI.2025.3530156

Iqbal, E., Niaz, A., Memon, A. A., Asim, U., & Choi, K. N. (2020). Saliency-Driven Active Contour Model for image segmentation. *IEEE Access*, *8*, 208978–208991. https://doi.org/10.1109/ACCESS.2020.3038945

Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step data normalization approach for improving classification accuracy in the medical diagnosis domain. *Mathematics*, *10*(11), 1–18. https://doi.org/10.3390/math10111942

Kalaiyarasi, M., Dhanasekar, R., Sakthiya Ram, S., & Vaishnavi, P. (2020). Classification of benign or malignant tumor using machine learning. *IOP Conference Series: Materials Science and Engineering*, *995*(1), 1–12. https://doi.org/10.1088/1757-899X/995/1/012028

Le, T. T., Fu, W., & Moore, J. H. (2020). Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*, *36*(1), 250–256. https://doi.org/10.1093/bioinformatics/btz470

Lee, C. S., Moy, L., Hughes, D., Golden, D., Bhargavan-Chatfield, M., Hemingway, J., Geras, A., Duszak, R., & Rosenkrantz, A. B. (2021). Radiologist characteristics associated with interpretive performance of screening mammography: A national mammography database (NMD) study. *Radiology*, *300*(3), 518–528. https://doi.org/10.1148/radiol.2021204379

Mahmood, T., Saba, T., Rehman, A., & Alamri, F. S. (2024). Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert Systems with Applications*, *249*(PC), 123747. https://doi.org/10.1016/j.eswa.2024.123747

Moore, J. H., Ribeiro, P. H., Matsumoto, N., & Saini, A. K. (2023). *Machine Learning—Automated Machine Learning (AutoML) for disease prediction BT - clinical applications of Artificial Intelligence in real-world data* (F. W. Asselbergs, S. Denaxas,

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 21 (June 2025) PP. 01-15
DOI 10.35631/IJIREV.721001

D. L. Oberski, & J. H. Moore, Eds.; pp. 161–173). Springer International Publishing. https://doi.org/10.1007/978-3-031-36678-9_10

Nguyen-Tat, T. B., Hung, T. Q., Nam, P. T., & Ngo, V. M. (2025). Evaluating pre-processing and deep learning methods in medical imaging: Combined effectiveness across multiple modalities. *Alexandria Engineering Journal*, *119*(October 2024), 558–586. https://doi.org/10.1016/j.aej.2025.01.090

Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a tree-based pipeline optimization tool for automating data science. *Proceedings of the Genetic and Evolutionary Computation Conference*, 485–492. https://doi.org/10.1145/2908812.2908918

Radzi, S. F. M., Karim, M. K. A., Saripan, M. I., Rahman, M. A. A., Isa, I. N. C., & Ibahim, M. J. (2021). Hyperparameter tuning and pipeline optimization via grid search method and tree-based autoML in breast cancer prediction. *Journal of Personalized Medicine*, *11*(10), 1–17. https://doi.org/10.3390/jpm11100978

Radzi, S. Fairuz. M., Muhammad Khalis, A. K., Saripan, M. I., Abd Rahman, M. A., Osman, N. H., Dalah, E. Z., & Noramaliza, M. N. (2020). Impact of image contrast enhancement on stability of radiomics feature quantification on a 2D mammogram radiograph. *IEEE Access*, *8*, 127720–127731. https://doi.org/10.1109/ACCESS.2020.3008927

Rashed, A. E. E., Elmorsy, A. M., & A. E. M. (2023). Comparative evaluation of automated machine learning techniques for breast cancer diagnosis. *Biomedical Signal Processing and Control*, *86*(105016), 1–18. https://doi.org/10.1016/j.bspc.2023.105016

Schaffter, T., Buist, D. S. M., Lee, C. I., Nikulin, Y., Ribli, D., Guan, Y., Lotter, W., Jie, Z., Du, H., Wang, S., Feng, J., Feng, M., Kim, H. E., Albiol, F., Albiol, A., Morrell, S., Wojna, Z., Ahsen, M. E., Asif, U., … Jung, H. (2020). Evaluation of combined Artificial Intelligence and radiologist assessment to interpret screening mammograms. *JAMA Network Open*, *3*(3), 1–15. https://doi.org/10.1001/jamanetworkopen.2020.0265

Smolarz, B., Zadrożna Nowak, A., & Romanowicz, H. (2022). Breast Cancer—Epidemiology, classification, pathogenesis and treatment (review of literature). In *Cancers* (Vol. 14, Issue 10, pp. 1–27). MDPI. https://doi.org/10.3390/cancers14102569

Srivastava, D., Rajitha, B., Agarwal, S., & Singh, S. (2020). Pattern-based image retrieval using GLCM. *Neural Computing and Applications*, *32*(15), 10819–10832. https://doi.org/10.1007/s00521-018-3611-1

Tagliafico, A. S., Piana, M., Schenone, D., Lai, R., Massone, A. M., & Houssami, N. (2020). Overview of radiomics in breast cancer diagnosis and prognostication. *Breast*, *49*, 74–80. https://doi.org/10.1016/j.breast.2019.10.018

Vrigazova, B. P. (2020). Detection of malignant and benign breast cancer using the ANOVA-BOOTSTRAP-SVM. *Journal of Data and Information Science*, *5*(2), 62–75. https://doi.org/10.2478/jdis-2020-0012

Wang, G., Shi, D., Guo, Q., Zhang, H., Wang, S., & Ren, K. (2022). Radiomics based on digital mammography helps to identify mammographic masses suspicious for cancer. *Frontiers in Oncology*, *12*, 1–10. https://doi.org/10.3389/fonc.2022.843436

Yuan, H., Yu, K., Xie, F., Liu, M., & Sun, S. (2024). Automated machine learning with interpretation: A systematic review of methodologies and applications in healthcare. *Medicine Advances*, *2*(3), 205–237. https://doi.org/10.1002/med4.75

Zhou, Y., Wei, J., Wu, D., & Zhang, Y. (2022). Generating full-field digital mammogram from digitized screen-film mammogram for breast cancer screening with high-resolution generative adversarial network. *Frontiers in Oncology*, *12*(April), 1–9. https://doi.org/10.3389/fonc.2022.868257