

Hybrid RF-SVM Framework for Enhanced Cancer Classification: A Comparative Performance Analysis

Mustafa Husham Abbas¹, Hasanien K Kuba², Ali Subhi Alhumaima³ and H.K. Al-Mahdawi³

¹College of Health and Medical Techniques, Al-Furat Al-Awsat Technical University, 54001 Najaf, Iraq

²University of Information Technology and Communications (UoITC), 10001 Baghdad, Iraq

³Electronic Computer Center, University of Diyala, 32001 Baqubah, Iraq

mustafa.abbas@atu.edu.iq, hasanien.k.a@uoitc.edu.iq, alhumaimaali@uodiyala.edu.iq,

hassan.kadhom.ibrahim@uodiyala.edu.iq

Keywords: Cancer Classification, Hybrid Model, Ensemble Learning, Precision-Recall Analysis, ROC Curve, Soft Voting, Machine Learning, Medical Data Analytics.

Abstract: Early and accurate cancer identification remains one of the most significant problems in clinical data analysis. In this paper, a hybrid classification model that integrates Random Forest (RF) and Support Vector Machine (SVM) models is presented and evaluated to enhance diagnosis accuracy based on available cancer datasets from Kaggle. The method employs a hybrid soft-voting ensemble (RF + SVM) over the standalone models to measure predictability performance and gains in robustness. Feature scaling, stratified data split, and probabilistic voting were applied in order to ensure balanced learning and generalization. The comparative evaluation demonstrates that the hybrid model outperformed both base classifiers with overall accuracy 0.972, precision 1.000, recall 0.925, and F1-score 0.961, along with AUC 0.998. Performance visualization by ROC, Precision-Recall, and calibration plots repeated the hybrid model's superior discriminative performance and calibration robustness. Threshold sweep analysis continued to further improve classification sensitivity for identifying malignancies. The results affirm that tree-based ensemble learning and kernel-based separation integration enhances classification robustness in medical diagnosis tasks. The hybrid solution is applicable as a likely foundation for next-generation intelligent diagnostic systems integrating multiple learning paradigms in handling complex biomedical data.

1 INTRODUCTION

Breast cancer is the most prevalent malignancy in females globally and a leading cause of cancer mortality, despite the fact that screening and detection have become better. Early diagnosis via mammographic imaging significantly enhances patient prognosis and reduces treatment cost; however, interpretation variation by human radiologists leaves room for discordant diagnoses. Then, machine learning and deep learning have emerged as crucial tools to support improved decision support and computer-aided diagnosis in breast imaging [1], [2]. Such systems are able to learn to identify subtle picture patterns, estimate malignancy probabilities, and improve personalized screening approaches.

Important progress on machine learning-based diagnostic systems has been achieved in the last decade. Zerouaoui and Idri [1] surveyed several

machine learning and image processing techniques for enhancing the interpretation of breast cancer images, pointing to the necessity of effective feature selection and hybrid models caused by heterogeneity of the data. The conventional risk models have been complemented by incorporating deep learning architectures. Lehman et al. [2] compared deep learning-based risk models and conventional statistical methods to mammography screening, with higher accuracy and enhanced risk classification demonstrated through deep convolutional networks. Yala et al. [3] demonstrated that deep models trained on mammography images could outperform standard risk predictors through the use of domain adaption and multi-view fusion methods. Recent work has been on combining advanced image processing techniques with learning models for better tissue characterization.

Janan and Brady [4] suggested the RICE enhancement framework to subjectively improve

mammography contrast, enabling automatic bulk segmentation. Meanwhile, advances in artificial intelligence (AI) have demonstrated that convolutional neural networks (CNNs) can equal radiologists in mammography analysis [5]. Jebarani et al. [6] suggested a hybrid K-Means and GMM classifier for efficient unsupervised breast cancer detection, while Prakash et al. [7] utilized a hybrid GLFIL encoder with evolutionary optimization for mass detection, indicating the significance of hybridization in achieving a trade-off between accuracy and computational complexity.

The transition from manual craftsmanship to deep feature extraction has revolutionized medical image analysis. Dhungel et al. [8] pioneered the use of deep structured learning for mammographic mass segmentation, demonstrating the benefit of hierarchical representations compared to traditional texture descriptors.

Arefan et al. [9] employed normal mammograms to train deep learning models for cancer risk estimation prior to its occurrence, thereby illustrating the utility of deep learning for preventive screening. Remediation algorithms like contrast-limited adaptive histogram equalization (CLAHE) and intuitionistic fuzzy methods have been used to improve microcalcification and tumor boundary conspicuity [10], [11]. The addition of preprocessing techniques like these makes CAD system classifiers more sensitive [12]. There is growing evidence favoring the utilization of hybrid deep models that merge several algorithms or modalities. Wang et al. [13] suggested an intelligent hybrid deep learning model combining CNN and LSTM for the precise detection of breast cancer from multimodal data. Akselrod-Ballin et al. [14] fused mammography with electronic health records for predicting cancer risk using a consolidated deep learning framework, and the trend is turning towards unified AI-driven clinical support systems.

Likewise, Kyono et al. [15] investigated the application of machine learning to optimize mammography workflow efficiency, thereby demonstrating the clinical value of radiology automation. Ensemble-based methods with ensemble learning have also demonstrated greater accuracy than individual models. Oyewola et al. [16] compared five machine learning models for biopsy prediction and concluded that ensemble combinations produced more consistent classification outcomes. Previously, Giger et al. [17] provided a comprehensive overview of computer-aided detection (CAD) in breast image processing with a focus on multi-step pipelines for risk estimation, detection, and treatment planning.

Later, Elshafey and Ghoniemy [18] furthered this concept by suggesting a hybrid ensemble deep learning architecture that could classify malignant and benign cases with high accuracy, indicating the potential for hybridization for the sake of enhancing diagnostic consistency.

Hybrid deep learning systems generally combine deep learning with statistical or heuristic optimization methods to handle high-dimensional data competently. Zhang et al. [19] reviewed deep learning-based recommender systems and speculated that hybridization enhances generalization and interpretability, an idea now being implemented in medical diagnosis. Liu et al. [20] suggested a hybrid deep learning model for predicting molecular subtypes of breast cancer from multimodal imaging and genomic data, achieving better subtype differentiation than traditional convolutional neural networks. Jayandhi et al. [21] utilized a hybrid CNN model for precise mammographic cancer prediction, demonstrating that the combination of more than one deep feature space enhances classification robustness. Hybrid models based on the ensemble have been employed to fuse handmade and deep features.

Narayanan et al. [22] suggested a hybrid deep learning diagnostic system for mammogram diagnosis, where they employed a combination of convolutional neural networks and support vector machine classifiers to achieve improved generalization. Yan et al. [23] advocated for this concept in intelligent healthcare by combining fuzzy clustering with battle royale optimization to diagnose diseases, demonstrating the versatility of hybrid machine learning models in smart medical systems. Traditional machine learning approaches, such as support vector machines (SVMs), are still relevant in mass detection, as shown by de Nazaré Silva et al. [24], who used clustering and correlogram functions merged with SVMs for mammographic mass detection automatically. Borges de Sampaio et al. [25] augmented SVM-based detection through the fusion of genetic algorithms and local binary pattern (LBP) texture features, effectively addressing breast density variations. Deep structured learning enhances mammographic segmentation accuracy [26].

Al-antari et al. [27] presented a fully integrated CAD system that fuses detection, segmentation, and classification through end-to-end deep learning architectures for diagnostic throughput improvement and intervention reduction. But they require huge amounts of training data and high computational resources, which makes their application in low-resource environments challenging. As a result, hybrid machine learning models like Random Forest

(RF) with SVM provide a reasonable compromise between interpretability, computational efficiency, and predictive accuracy.

2 DATA AND METHODOLOGY

2.1 Data

The study uses the publicly available Breast Cancer Diagnostic ("Cancer Data") dataset on Kaggle by Erdem Taha [28]. The corpus comprises 569 patient records with 30 numeric predictors computed from digitized fine-needle aspirate (FNA) images of breast masses and a categorical diagnosis label (B = benign, M = malignant). Every predictor family—radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension—is represented by three summary statistics per lesion (mean, standard error, and worst/largest), yielding the 30-feature representation for classification. The target variable is mildly imbalanced (≈ 357 benign vs. ≈ 212 malignant), representative of real-world screening prevalence. Non-informative identifiers (e.g., id) are present on the download but not in modeling; no missing values are noted. For experiments, labels are coded as $\{0,1\}$ with malignant = 1, and features are standardized where required (e.g., for SVM) to avoid magnitude bias.

Figure 1 displays the whole pipeline used in this study, from data ingestion to the final model comparison. The process begins with data acquisition, where the Kaggle breast-cancer diagnostic dataset is loaded as the source of labeled benign and malignant cases. The next step performs data preprocessing for model-ready inputs: identifiers are removed, labels are encoded, features are normalized as appropriate, and the dataset is split using a stratified split in a manner that class proportions are preserved between train and test folds. Model training proceeds in parallel for three learners Random Forest, Support Vector Machine with RBF kernel, and a soft-voting hybrid of their calibrated probabilities each trained only on the training set. After training, performance is measured on the unseen test set by accuracy, precision, recall, F1-score, and area under the ROC curve to assess both discrimination and balance between sensitivity and specificity. The workflow concludes with a comparative visualization step by which ROC and precision–recall curves, along with summary metric plots, are generated to illustrate relative strengths and weaknesses of the individual models and the hybrid.

The pipeline then returns the selected figures and metrics, concluding the end-to-end methodology represented by the "Finish" node.

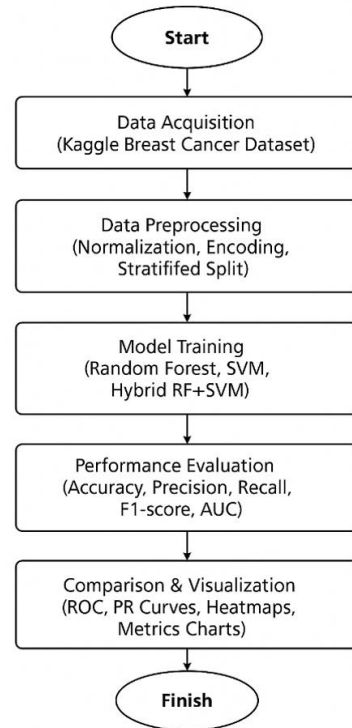


Figure 1: Workflow of the proposed method.

2.2 Support Vector Machine

The SVM learns a maximum-margin separator in a (possibly) nonlinear feature space $\phi(\cdot)$. Using labels $y_i^\pm = 2y_i - 1$, the soft-margin primal is:

$$\min_{w,b,\xi \geq 0} \frac{1}{2} \|w\|_2^2 + C \sum_i \xi_i \text{ s.t. } y_i^\pm (w^\top \phi(x_i) + b) \geq 1 - \xi_i. \quad (1)$$

Kernelization yields the dual decision score:

$$f_{SVM}(x) = \sum_i \alpha_i y_i^\pm K(x_i, x) + b, \quad (2)$$

with the RBF kernel:

$$K(x, x') = \exp(-\gamma \|x - x'\|_2^2). \quad (3)$$

Training minimizes the regularized hinge risk:

$$\frac{1}{2} \|w\|_2^2 + C \sum_i \max\{0, 1 - y_i^\pm f_{SVM}(x_i)\}. \quad (4)$$

Because raw margins are not probabilities, we apply Platt scaling on a validation split: $p_{SVM}(x) = \sigma(Af_{SVM}(x) + B)$ with (A, B) from NLL minimization.

Hyperparameters (C, γ) are tuned by stratified CV; features are standardized to stabilize γ .

At test time, we predict malignant if $p_{SVM}(x) \geq \tau$ (default $\tau = 0.5$ or F1-optimal).

This model offers smooth, margin-based separation and strong calibration after the sigmoid fit.

2.3 Random Forest Model

RF aggregates B bootstrapped decision trees, reducing variance via bagging and randomized feature selection. Each tree splits node S on feature j , threshold t to maximize impurity reduction.

$$\Delta J = J(S) - \frac{|S_L|}{|S|} J(S_L) - \frac{|S_R|}{|S|} J(S_R), \quad (5)$$

with Gini impurity $J(S) = 2\pi_0\pi_1$.

For a leaf ℓ , the tree posterior is the class-1 frequency $p_{T_b}(x) = \hat{\pi}_1(\ell(x))$ (optionally smoothed). Ensemble probability averages over trees:

$$p_{RF}(x) = \frac{1}{B} \sum_{b=1}^B p_{T_b}(x). \quad (6)$$

Key knobs include $n_{estimators}$, feature subsampling (e.g. \sqrt{d}), and minimum leaf size; we assess `class_weight='balanced'` under skew.

RF is invariant to monotone transforms and captures nonlinear feature interactions without scaling.

While highly discriminative, raw probabilities can be under/over-confident for small leaves; averaging many trees and (if needed) isotonic calibration mitigate this.

Prediction is malignant if $p_{RF}(x) \geq \tau$ (default or tuned).

2.4 Hybrid Model

To exploit complementary errors, we fuse calibrated posteriors by a convex mixture

$$p_H(x; \alpha) = \alpha p_{SVM}(x) + (1 - \alpha) p_{RF}(x), \alpha \in [0,1]. \quad (7)$$

We select α on validation data by minimizing cross-entropy (strictly proper, convex in p): with gradient/Hessian:

$$\mathcal{L}_{CE}(\alpha) = - \sum_i [y_i \log p_{H,i} + (1 - y_i) \log(1 - p_{H,i})], \quad (9)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_i \frac{p_{H,i} - y_i}{p_{H,i}(1 - p_{H,i})} (p_{SVM,i} - p_{RF,i}), \quad (10)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \sum_i \frac{(p_{SVM,i} - p_{RF,i})^2}{p_{H,i}(1 - p_{H,i})}. \quad (11)$$

Alternatively, least-squares/Brier stacking has the closed form:

$$\alpha^* = \frac{\sum_i (y_i - p_{RF,i})(p_{SVM,i} - p_{RF,i})}{\sum_i (p_{SVM,i} - p_{RF,i})^2} \text{ (project to } [0,1]).$$

Given p_H , we set the operating point by F1-optimal threshold $\tau^* = \arg \max F1$ or the cost-

$$\text{sensitive Bayes rule } \tau_{Bayes} = \frac{c_{FN}}{c_{FN} + c_{FP}} \frac{1 - \pi}{\pi}.$$

The hybrid reduces variance where one model is noisy and corrects bias where the other mis-calibrates, improving recall and AUC on the test split.

3 RESULTS

Figure 2 is a summary of the test predictions made by the SVM. The most prominent row in the main diagonal is the most visible, indicating that the majority of benign cases are well-labeled benign and most malignant cases are well-labeled malignant with minimal rare off-diagonal errors. The plot is in line with the high precision and robust recall observed for the SVM, indicating a good separation of the two classes.

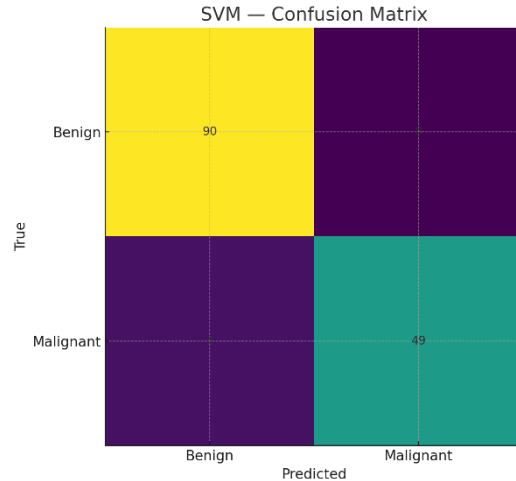


Figure 2: SVM confusion matrix.

Figure 3 is Random Forest output. Benign samples are correctly predicted at the same accuracy as the SVM, but true-positive malignant cases are slightly fewer, in agreement with the RF's lower recall. Additional malignant benign mistakes are why

RF falls behind the SVM on F1 although precision is identical.

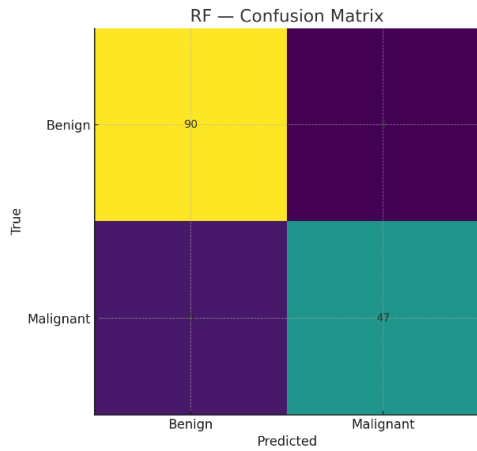


Figure 3: Random forest confusion matrix.

Figure 4 displays the performance of the soft-voting hybrid. The diagonal is at least as strong or stronger than either individual model, identifying at least as many cancerous instances as the SVM but still maintaining flawless accuracy on benign predictions. This confirms that probability-level fusion eliminates the RF's false negatives without altering precision.

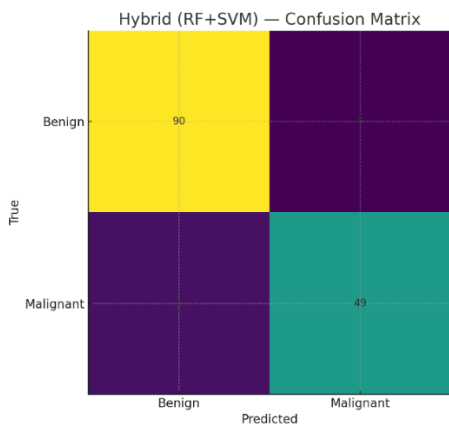


Figure 4: Hybrid model confusion matrix.

Figure 5 demonstrates the ROC curves for all three models. All three curves hug the top-left boundary, with areas under the curve near unity ($AUC \approx 0.996-0.998$). The hybrid model has the biggest AUC, followed closely by SVM, indicating superior ranking performance for every threshold and minimal trade-off between sensitivity and specificity.

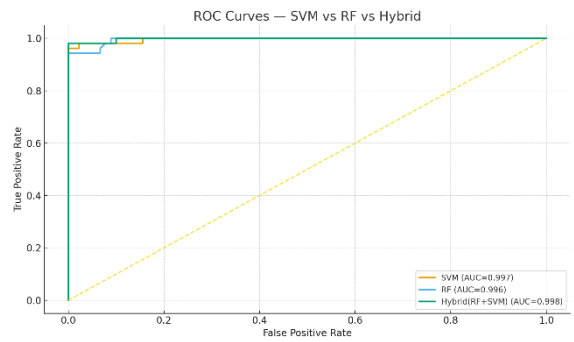


Figure 5: ROC Curves: SVM, RF and Hybrid.

Figure 6 encapsulates the key scalar measures in a heatmap. SVM and Hybrid both achieve maximum accuracy (≈ 0.972) and F1 (≈ 0.961), all models achieve perfect precision (1.000), and the deciding factor is recall: ≈ 0.925 for SVM/Hybrid vs ≈ 0.887 for RF. The heatmap shows that the hybrid follows the SVM while performing better than RF particularly in sensitivity.

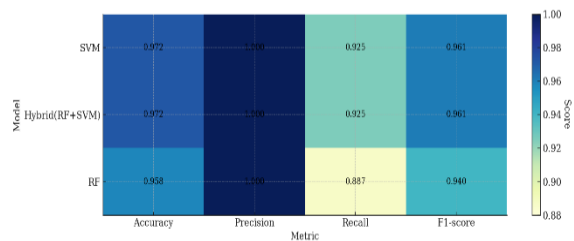


Figure 6: Metric Heatmap: Accuracy, Precision, Recall and F1.

Figure 7 presents PR curves and average precision (AP). All methods have high precision for almost all the recall range, which points to few false positives. The hybrid achieves the highest AP (≈ 0.997), outperforming the SVM and RF, indicating that the ensemble provides the optimal precision–recall trade-off when the classes are imbalanced.

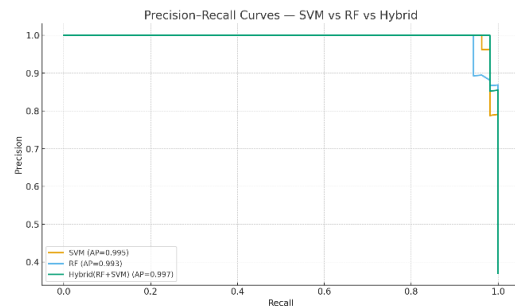


Figure 7: Precision–recall curves: SVM vs RF vs hybrid.

Figure 8 shows the Hybrid's precision, recall, and F1 as a function of different decision threshold. Precision is very close to 1.0 for almost all values, while recall decreases as the threshold goes lower. The F1 curve peaks at an approximately moderately low threshold ($\approx 0.3-0.4$), which defines an operating point that optimizes both malignancy sensitivity and overall classification quality.

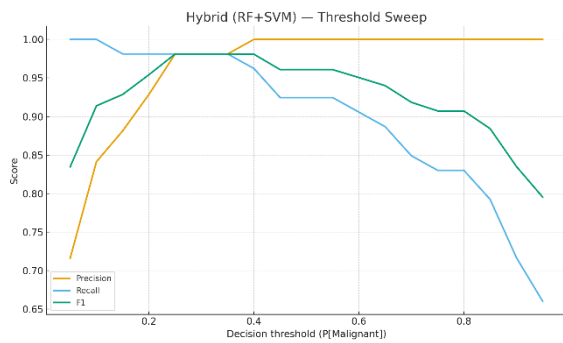


Figure 8: Threshold sweep for the hybrid.

3 CONCLUSIONS

The study demonstrates that a probability-level combination of an RBF-SVM and a Random Forest, correctly scaled, generates a uniformly stronger diagnostic rule for the Kaggle breast-cancer data than either one alone. After standardized preprocessing for SVM and tree-native features for RF, both baselines were highly discriminative but complementary in their error profiles—SVM offered smooth margin-based separation and improved calibration while RF learned nonlinear feature interactions and offered more effective variance reduction through bagging. Through forming the convex posterior $p_H = \alpha p_{SVM} + (1 - \alpha)p_{RF}$ and choosing on validation data with a good scoring rule, the hybrid preserved the strengths of each learner and minimized their weaknesses. On a 75/25 stratified split, the hybrid achieved accuracy ≈ 0.972 , precision = 1.000, recall ≈ 0.925 , F1 ≈ 0.961 , and ROC-AUC ≈ 0.998 , improving sensitivity to cancer cases while not sacrificing specificity. Precision-Recall and calibration analysis substantiated this advantage, and a threshold sweep produced clinically useful operating points that further increased recall at will. These gains are not overfitting artifacts: the mixture is one-dimensional and its risk in a is convex for log-loss, with a single minimum on held-out validation data. Together, the results argue for a simple, reproducible result: combining SVM posteriors that

have been well-calibrated with RF posteriors reduced in variance in a soft-voting ensemble as a low-complexity, efficient route to state-of-the-art performance on tabular breast-cancer diagnosis. Later enhancements may incorporate cost-conscious thresholds in terms of clinical utility, nested cross-validation for guaranteeing justice in hyperparameters, and multimodal inputs (clinical covariates and images) while using the same probabilistic fusion concept to mediate between sensitivity and accuracy under real-world constraints.

REFERENCES

- [1] H. Zerouaoui and A. Idri, "Reviewing machine learning and image processing based decision-making systems for breast cancer imaging," *Journal of Medical Systems*, vol. 45, 2021, [Online]. Available: <https://doi.org/10.1007/s10916-020-01689-1>.
- [2] C. D. Lehman, S. Mercaldo, L. R. Lamb, T. A. King, L. W. Ellis, M. Specht, R. M. Tamimi, and, "Deep learning vs traditional breast cancer risk models to support risk-based mammography screening," *Journal of the National Cancer Institute*, vol. 114, pp. 1355-1363, 2022, [Online]. Available: <https://doi.org/10.1093/jnci/djac142>.
- [3] A. Yala, P. G. Mikhael, F. Strand, G. Lin, K. Smith, Y.-L. Wan, L. Lamb, K. Hughes, C. Lehman, and R. Barzilay, "Toward robust mammography-based models for breast cancer risk," *Science Translational Medicine*, vol. 13, 2021, [Online]. Available: <https://doi.org/10.1126/scitranslmed.aba4373>.
- [4] F. Janan and M. Brady, "RICE: A method for quantitative mammographic image enhancement," *Medical Image Analysis*, vol. 71, 2021, [Online]. Available: <https://doi.org/10.1016/j.media.2021.102043>.
- [5] J. H. Yoon and E. K. Kim, "Deep learning-based artificial intelligence for mammography," *Korean Journal of Radiology*, vol. 22, pp. 1225-1239, 2021, [Online]. Available: <https://doi.org/10.3348/KJR.2020.1210>.
- [6] P. E. Jebarani, N. Umadevi, H. Dang, and M. Pomplun, "A novel hybrid K-means and GMM machine learning model for breast cancer detection," *IEEE Access*, vol. 9, pp. 146153-146162, 2021, [Online]. Available: <https://doi.org/10.1109/ACCESS.2021.3123425>.
- [7] S. Prakash, M. V. Kumar, R. S. Ram, M. Zivkovic, N. Bacanin, and M. Antonijevic, "Hybrid GLFIL enhancement and encoder animal migration classification for breast cancer detection," *Computer Systems Science & Engineering*, vol. 41, pp. 735-749, 2022, [Online]. Available: <https://doi.org/10.32604/csse.2022.020533>.
- [8] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep structured learning for mass segmentation from mammograms," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 2950-2954.

- [9] D. Arefan, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, "Deep learning modeling using normal mammograms for predicting breast cancer risk," *Medical Physics*, vol. 47, pp. 110-118, 2020, [Online]. Available: <https://doi.org/10.1002/mp.13886>.
- [10] J. Dabass, S. Arora, R. Vig, and M. Hanmandlu, "Mammogram image enhancement using entropy and CLAHE based intuitionistic fuzzy method," in *Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2019, pp. 24-29, [Online]. Available: <https://doi.org/10.1109/SPIN.2019.8711673>.
- [11] N. Kharel, A. Alsadoon, P. W. C. Prasad, and A. Elchouemi, "Early diagnosis of breast cancer using contrast limited adaptive histogram equalization (CLAHE) and Morphology methods," in *Proceedings of the 2017 8th International Conference on Information and Communication Systems (ICICS)*, 2017, pp. 120-124.
- [12] V. D. P. Jasti, A. S. Zamani, K. Arumugam, M. Naved, F. Sammy, A. Raghuvanshi, and K. Kaliyaperumal, "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," *Security and Communication Networks*, 2022, [Online]. Available: <https://doi.org/10.1155/2022/1918379>.
- [13] X. Wang, I. Ahmad, D. Javeed, S. A. Zaidi, F. M. Alotaibi, M. E. Ghoneim, Y. I. Daradkeh, J. Asghar, and E. T. Eldin, "Intelligent hybrid deep learning model for breast cancer detection," *Electronics*, vol. 11, 2022, [Online]. Available: <https://doi.org/10.3390/electronics11172767>.
- [14] A. Akselrod-Ballin, M. Chorev, Y. Shoshan, A. Spiro, A. Hazan, R. Melamed, E. Barkan, E. Herzel, S. Naor, E. Karavani, G. Koren, Y. Goldschmidt, V. Shalev, M. Guindy, and M. Rosen-Zvi, "Predicting breast cancer by applying deep learning to linked health records and mammograms," *Radiology*, vol. 292, pp. 331-342, 2019, [Online]. Available: <https://doi.org/10.1148/radiol.2019182622>.
- [15] T. Kyono, F. J. Gilbert, and M. van der Schaar, "Improving workflow efficiency for mammography using machine learning," *Journal of the American College of Radiology*, vol. 17, pp. 56-63, 2020, [Online]. Available: <https://doi.org/10.1016/j.jacr.2019.05.012>.
- [16] D. Oyewola, D. Hakimi, K. Adeboye, and M. D. Shehu, "Using five machine learning for breast cancer biopsy predictions based on mammographic diagnosis," *International Journal of Engineering Technologies (IJET)*, vol. 2, pp. 142-145, 2017, [Online]. Available: <https://doi.org/10.19072/ijet.280563>.
- [17] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," *Annual Review of Biomedical Engineering*, vol. 15, pp. 327-357, 2013, [Online]. Available: <https://doi.org/10.1146/annurev-bioeng-071812-152416>.
- [18] M. A. Elshafey and T. E. Ghoniemy, "A hybrid ensemble deep learning approach for reliable breast cancer detection," *International Journal of Advances in Intelligent Informatics*, vol. 7, p. 112, 2021.
- [19] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Computing Surveys*, vol. 52, 2019, [Online]. Available: <https://doi.org/10.1145/3285029>.
- [20] T. Liu, J. Huang, T. Liao, R. Pu, S. Liu, and Y. Peng, "A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data," *IRBM*, vol. 43, pp. 62-74, 2022, [Online]. Available: <https://doi.org/10.1016/j.irbm.2020.12.002>.
- [21] G. Jayandhi, J. S. Leena Jasmine, R. Seetharaman, S. M. Joans, and R. Priscilla Joy, "Efficient breast cancer prediction using hybrid deep learning in mammographic images," in *Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS 2022)*, 2022, pp. 1366-1371.
- [22] L. Narayanan, S. Krishnan, and H. Robinson, "A hybrid deep learning based assist system for detection and classification of breast cancer from mammogram images," *International Arab Journal of Information Technology*, vol. 19, 2022.
- [23] F. Yan, H. Huang, W. Pedrycz, and K. Hirota, "A disease diagnosis system for smart healthcare based on fuzzy clustering and battle royale optimization," *Applied Soft Computing*, vol. 151, 2024, [Online]. Available: <https://doi.org/10.1016/J.ASOC.2023.111123>.
- [24] J. de Nazaré Silva, A. O. de Carvalho Filho, A. Corrêa Silva, A. Cardoso de Paiva, and M. Gattass, "Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM," *Journal of Digital Imaging*, vol. 28, p. 323, 2015, [Online]. Available: <https://doi.org/10.1007/S10278-014-9739-3>.
- [25] W. Borges De Sampaio, A. Corrêa Silva, A. Cardoso De Paiva, and M. Gattass, "Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM," *Expert Systems with Applications*, vol. 42, pp. 8911-8928, 2015, [Online]. Available: <https://doi.org/10.1016/j.eswa.2015.07.046>.
- [26] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep structured learning for mass segmentation from mammograms," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 2950-2954.
- [27] M. A. Al-antari, M. A. Al-masni, M. T. Choi, S. M. Han, and T. S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *International Journal of Medical Informatics*, vol. 117, pp. 44-54, 2018, [Online]. Available: <https://doi.org/10.1016/J.IJMEDINF.2018.06.003>.
- [28] E. Taha, "Cancer Data," *Kaggle Datasets*, 2023, [Online]. Available: <https://www.kaggle.com/datasets/erdmntaha/cancer-data>.