Normal and Abnormal Classification of Medical Thermal Images by DNN and RF Algorithms

Zainab Mohammed Essa[1], Doaa Mohsin Abd Ali Afraji1, Ziad M. Abood[2], Aya T. Hussein[3] and Aula T. Hussein[3]
[1]*Department of Computer Science, College of Education, Mustansiriyah University, 10052 Baghdad, Iraq*
[2]*Department of Physics, College of Education, Mustansiriyah University, 10052 Baghdad, Iraq*
[3]*College of Dentistry, University of Diyala, 32001 Diyala, Iraq*
dr.ziadmabood@uomustansiriyah.edu.iq

Keywords: AI Algorithms, Classification, Feature Extraction, Breast Thermal Images, DNN, RF.

Abstract: Deep learning (DL) and machine learning (ML) techniques are used in AI-based "normal" vs. "abnormal" data classification, especially in healthcare and industrial applications, to differentiate between standard, healthy, or routine states and deviations that indicate illness, injury, or malfunctioning operations. This technique is commonly utilized to increase productivity, decrease human error, and speed up decision-making in signal processing, behavior monitoring, and imaging. One of the most recent developments in medical imaging is thermal imaging of breast cancers. Conventional techniques make it more difficult for doctors to diagnose patients quickly and accurately since they take longer to evaluate every image. Research on computer-aided techniques for rapidly and precisely diagnosing all produced pictures has been prompted by this. It is time-consuming and prone to inaccuracy to manually detect and categorize breast cancers. The system makes use of artificial intelligence algorithms to compute certain metrics for a given region or to identify the best configuration of picture attributes for categorization. This study focused on the use of machine learning in the diagnosis and classification of breast tumors using thermal imaging. The results of the current research indicate that the algorithms used in the proposed classification system perform well based on image quality metrics. The accuracy of classifying benign from malignant breast tumors reached 90.25% and 86.65%, using DNN and RF algorithms respectively.

1 INTRODUCTION

The automated diagnosis of disorders affecting different organs, such as the skin, brain, and breast, has advanced significantly during the last 20 years in medical imaging technology. The most prevalent of these illnesses in women is breast cancer, which affects the health of a sizable portion of women globally. It is regarded as one of the main reasons why women die [1].

The risk of mortality can be considerably decreased by early diagnosis of breast cancers. Manual evaluation is a tedious and time-consuming procedure when compared to computer-based techniques for determining tumor type and size. Breast illnesses are categorized and classified using both deep learning and traditional methods. While deep learning (artificial intelligence) algorithms can extract valuable features throughout the analysis process based on medical pictures (particularly thermal imaging of the breast), traditional methods rely on manually picking predetermined criteria. Accurate early identification of breast cancers is still difficult, despite significant efforts in this area [2].

Scientific computing makes extensive use of deep learning, a subfield of artificial neural networks. Numerous industries employ deep learning techniques to address challenging issues. Different types of neural networks are used by all deep learning techniques. It falls within the category of machine learning. Numerous fields, including computer vision, medical applications, picture categorization, multimedia concept retrieval, and medical information processing, have made use of it. Deep learning (DL) is one of the most popular kinds of algorithms [3].

Supervised segmentation techniques use training samples to leverage prior information about the image processing problem. One type of supervised model that became popular in the 1990s is atlas-based segmentation methods. These techniques, including statistical shape models and probabilistic atlases, are more accurate than unsupervised models at capturing organ form. Other supervised segmentation methods that have been extensively studied over the past ten years include deep neural networks (DNN), support vector machines (SVM), random forests (RF), and K-nearest neighbor clustering. Nevertheless, these methods have not been very successful in identifying hazy organ boundaries in radiological pictures [4].

Thermography, commonly referred to as Digital Infrared Thermal Imaging (DITI), is a non-invasive, radiation-free diagnostic method that uses high-sensitivity cameras to record skin surface temperatures. By detecting physiological changes through the detection of heat patterns associated with blood flow and inflammation, it is commonly used as an adjunct to mammography for early breast cancer screening, although it is not a stand-alone replacement. [5]

Thermography, a useful method for the temperature of the breast region projected from the skin surface to the surroundings is measured for the early detection of anomalies in the female breast.

A thermogram is a temperature picture created by arranging the temperatures in a two-dimensional array This is the result of the radiation intensity of the region being studied being captured by a thermographic camera and transformed by its system for processing signals (Fig. 1). Ultimately, every thermogram is arranged into data sets [6].

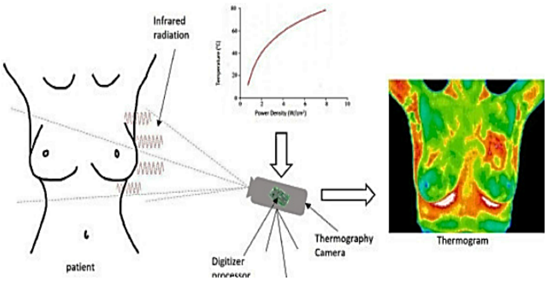

Figure 1: Image capture and radiation intensity transformation in the research region [6].

2 ARTIFICIAL INTELLIGENCE (AI)

Artificial intelligence (AI) and machine learning (ML) have made rapid progress in recent years. ML and AI have helped medical image processing, computer-aided diagnosis, picture interpretation, image fusion, image registration, image segmentation, image retrieval, and analysis. ML effectively and efficiently expresses information by extracting information from pictures [7]. AI and ML can more rapidly and effectively detect, predict, and prevent illnesses. There are several applications for machine learning and artificial intelligence (AI). In particular, these frameworks are crucial to healthcare, especially when it comes to medical diagnosis. A medical diagnosis is a method of classification. A physicist has to go over a series of symptoms and indicators [8]. before making a diagnosis that would make their line of work more difficult (Fig. 2).

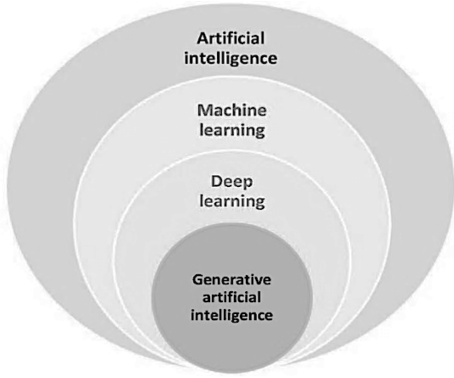

Figure 2: Al, ML and DL.

3 RELATED WORKS

Recent research has demonstrated how the application of deep learning techniques in the categorization of skin lesions-in this example, melanoma detection-has transformed our lives. Wu et al. [9] went on to further summarize recent developments in deep learning that can be applied to the classification of skin cancer. They showed that transfer learning and fine-tuning of pretrained models resulted in a significant improvement in performance, especially when working with small or diverse datasets. Additionally, their findings demonstrated the significance of standardized assessment metrics including F1-score, balanced accuracy, and AUC in order to enable reliable cross-study comparisons.

Convolutional neural networks (CNNs) trained on dermoscopic images have been shown by Esteva et al. [10] to perform somewhat like board-certified dermatologists, with sensitivities in a clinically relevant range and accuracy rates comparable to reports from non-certified experts. In order to provide a baseline for further study, our work showed the possibility of applying data-driven methodologies in clinical picture interpretation [11].

The possibilities of human-computer cooperation in a deep learning study were examined by Tschandl et al. [12] in cooperation with specialists. Combining model prediction with human evaluation has been demonstrated to increase diagnostic accuracy and reliability when dermatologists consistently incorporated model predictions into clinical assessment.

According to Fuzhe and colleagues [12], the prevalence of chronic kidney disease (CKD) is increasing yearly. One of the sources for further therapy is the prediction of CKD, where machine learning techniques are becoming more and more important in medical diagnostics due to their high classification accuracy. The appropriate application of feature selection strategies to minimize the quantity of data is critical to the accuracy of classification algorithms. On the Internet of Medical Things platform, the Heterogeneous Modified Artificial Neural Network (HMANN) was created for the early detection, segmentation, and diagnosis of chronic renal failure.

Abood and Ali [13], A novel approach that has applications in medicine is thermal image modeling. The purpose of this study is to develop a method to categorize the breast's thermal areas as normal or pathological based on temperature fluctuations in the tissue that appear as distinct colors in thermal cameras that correlate to its bar. In order to determine the maximum and minimum of each characteristic for normal areas, the suggested technique first extracts features from known thermal pictures of a normal breast.

Ornek and Ceylan [14], study concentrate on either the statistical behavior of the thermograms' temperature distributions or simply the accurate classifications of the thermograms in order to ascertain the neonates' health condition. Nevertheless, categorization methods are never fully explainable. Physicians require explanations, particularly in medical research, in order to evaluate the potential outcomes of the choices. Introducing our latest research on Convolutional Neural Networks (CNNs) For the first time, Class Activation Maps (CAMs) have been used to display the health status of newborns. One of the pre-trained models, VGG16, has been chosen as a CNN model, and its final layers have been adjusted in accordance with CAMs. Train-validation accuracies exceeded 95% and test sensitivity-specificity ranged from 80.701% to 96.842% after the model was trained for 50 epochs. Our results show that the CNN mostly looks at the neck, armpit, and belly areas while learning the body's temperature distribution. The armpit and belly of healthy newborns are the areas of concentration, whereas the neck and abdomen of ill kids are. Therefore, we might say that CNN concentrates.

Nowakowski and Kaczmarek [15], the most recent developments in infrared thermal imaging techniques for use in medical diagnostics are examined. An overview of developments in infrared thermal imaging technology from 1960 to 2024 is provided. The primary focus is on recently developed artificial intelligence (AI) techniques for thermal image analysis. IR thermography is examined in light of cutting-edge machine learning techniques for better medical diagnosis and therapy. By denoising thermal images, using AI super-resolution algorithms, eliminating artifacts, object identification, face and distinctive feature localization, complicated matching of diagnostic symptoms, etc., the AI method seeks to improve picture quality.

4 METHODOLOGY

This study used methods based on popular supervised machine learning algorithms, DNN and RF, to categorize illnesses of a breast tumors. Both the training and testing stages made use of these methods.

4.1 Proposed Method

The proposed framework consists of a sequence of image acquisition, preprocessing, segmentation, feature extraction, and classification steps designed to improve the reliability and diagnostic accuracy of AI-based thermal image analysis for medical applications.

1) Data Acquisition & Protocol:
   Unlike standard photos, thermal images are sensitive to the environment.
   - Thermal Equilibration. Patients must rest in a temperature-controlled room (20°C–23°C) for 15 minutes so the body surface stabilizes.
   - Standardization. Images must be captured at a fixed distance and angle using a high-sensitivity infrared camera (at least 0.1°C thermal sensitivity).

2) Image Preprocessing:
   Raw thermal data frequently has little contrast or is noisy.
   - Using filters such as median or bilateral filters to eliminate "salt and pepper" noise without blurring edges is known as noise reduction. Normalization: To guarantee that

the AI handles every image same, independent of the camera's calibration, pixel values (temperature data) are converted into a standard range (0-1 or 0-255).

- Color mapping: Although many contemporary AI models favor the raw grayscale temperature values, this technique highlights hotspots by converting grayscale thermal data into pseudo-colors (such as Jet or Rainbow).

3) Segmentation & ROI Extraction:
The AI must only concentrate on the bodily part of interest (such as the breast, foot, or thyroid) and disregard the backdrop.
   - Manual or Auto-Cropping: Removing the background using methods such as GrabCut or Otsu's Thresholding.
   - DL Segmentation. This ensures that the categorization is based solely on pertinent tissue by automatically masking the particular organ using a U-Net architecture.

4) Feature Extraction:
The "Normal" and "Abnormal" differences are found here.
   - The "Gold Standard" of thermography is asymmetry analysis, which compares the left and right sides of the body. A discrepancy in temperature is frequently reported as abnormal.
   - Texture Features (GLCM) . Examining the "feel" of the heat distribution (homogeneity, contrast, correlation, and energy).
   - Deep Features. Making use of intricate, automatically learnt heat patterns that the human eye could overlook.

4.2 Classification (AI Model)

The processed data is fed into a classifier:

1) Supervised Learning: Training on labeled datasets (Normal vs. Abnormal) using:
   - RF: Good for smaller datasets with manual features.
   - DNN: implemented with fully connected hidden layers, ReLU activation, dropout regularization, and the Adam optimizer, generally the most accurate approach for large datasets, achieving over 95% accuracy in many studies; RF was configured as an ensemble of decision trees with Gini-impurity splitting. The exact layer counts, tree counts, and other hyperparameters used for the present dataset should be reported in a future revision for full reproducibility.
2) Anomaly Detection: If you have mostly "Normal" images, you can use Autoencoders. The model learns what "Normal" looks like and flags anything else as "Abnormal" [16].

Figure 3 shows examples of breast normal and abnormal medical cases [17], [18].

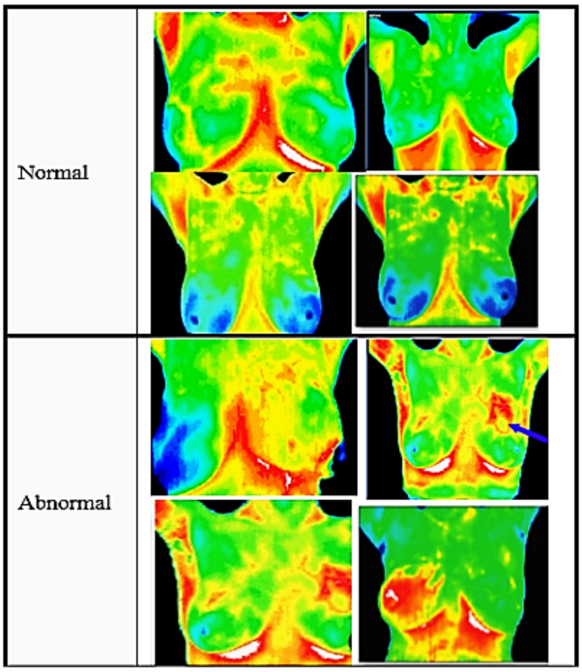

Figure 3: Represents thermal images of breast normal and abnormal medical cases [17], [18].

4.3 Performance Evaluation

The model is tested using:

- Sensitivity. How many sick patients did we correctly find? (Crucial in medicine to avoid missing cases).
- Specificity. How many healthy people were correctly identified? (To avoid unnecessary biopsies/anxiety).
- AUC-ROC Curve. To measure the overall diagnostic power of the algorithm.

Figure 4 show the main steps of present study.

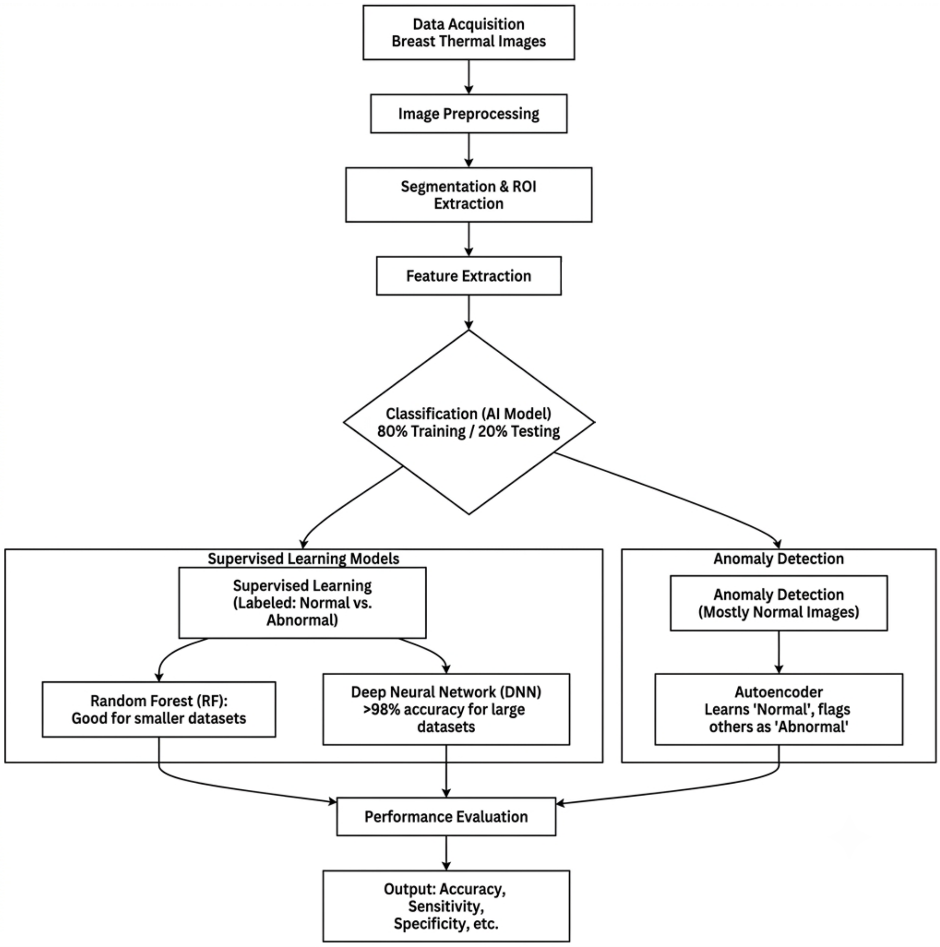

Figure 4: Block Diagram for study steps.

4.4 Feature Extraction

The following performance metrics, which included accuracy rate, precision, recall, and F1 measure for both normal and abnormal cases, demonstrated that the algorithms used in the suggested system for the classification of breast disease produced acceptable results based on the image quality metrics (Fig. 5):

- PSNR Peak Signal to Noise Ratio.
- NCC Normalized Cross-correlation.
- AD Average Difference.
- SSIM structural similarity index.
- SR-SIM Spectral Residuals Based Similarity Index.
- PCC Personal Correlation Coefficient.
- IQI Institutional Quality Index.

4.5 Step of Classification

Classification is the problem of grouping objects into distinct classes. The classification goal is to accurately anticipate the target class for each data sample. In machine learning, classification has been viewed as an example of supervised learning, which is learning in which a training set of correctly classified observations is provided. It is represented by two processes: learning and classification. constructing a classification model during the learning phase (Training Phase), when a range of methods have been used to learn the model's available training data and generate the classifier. To produce precise predictions, the model has to be trained [19]. A model is used in the classification step to classify the testing data and determine the classification rule precision, which is based on predetermined class labels [20]. One technique for assigning each item in a dataset to one of the predefined classes is dataset classification using a machine learning model [21]. Each of these classifiers builds a distinct model using the training dataset while operating in parallel [22]. The dataset of breast diseases is used to test classification systems

4.6 Evaluation Metrics

Performance metrics derived from the confusion matrix are commonly used to evaluate the effectiveness of machine learning models [23]. In this study, the confusion matrix was employed to assess the performance of the proposed detection model. The main diagonal represents correctly classified samples, whereas the off-diagonal elements correspond to misclassified cases. The confusion matrix used in this study is presented in Table 1.

The performance of the proposed method was evaluated using four widely adopted metrics: accuracy, precision, recall, and F1-score [24], [25]. Accuracy represents the overall proportion of correctly classified samples among all predictions. Precision measures the proportion of correctly identified positive samples among all samples predicted as positive. Recall, also referred to as sensitivity, indicates the proportion of actual positive samples that are correctly detected by the model. The F1-score combines precision and recall into a single performance measure and provides a balanced evaluation, particularly for datasets with imbalanced class distributions.

Table 1: Confusion matrix Prediction.

Actual class	Predicted class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

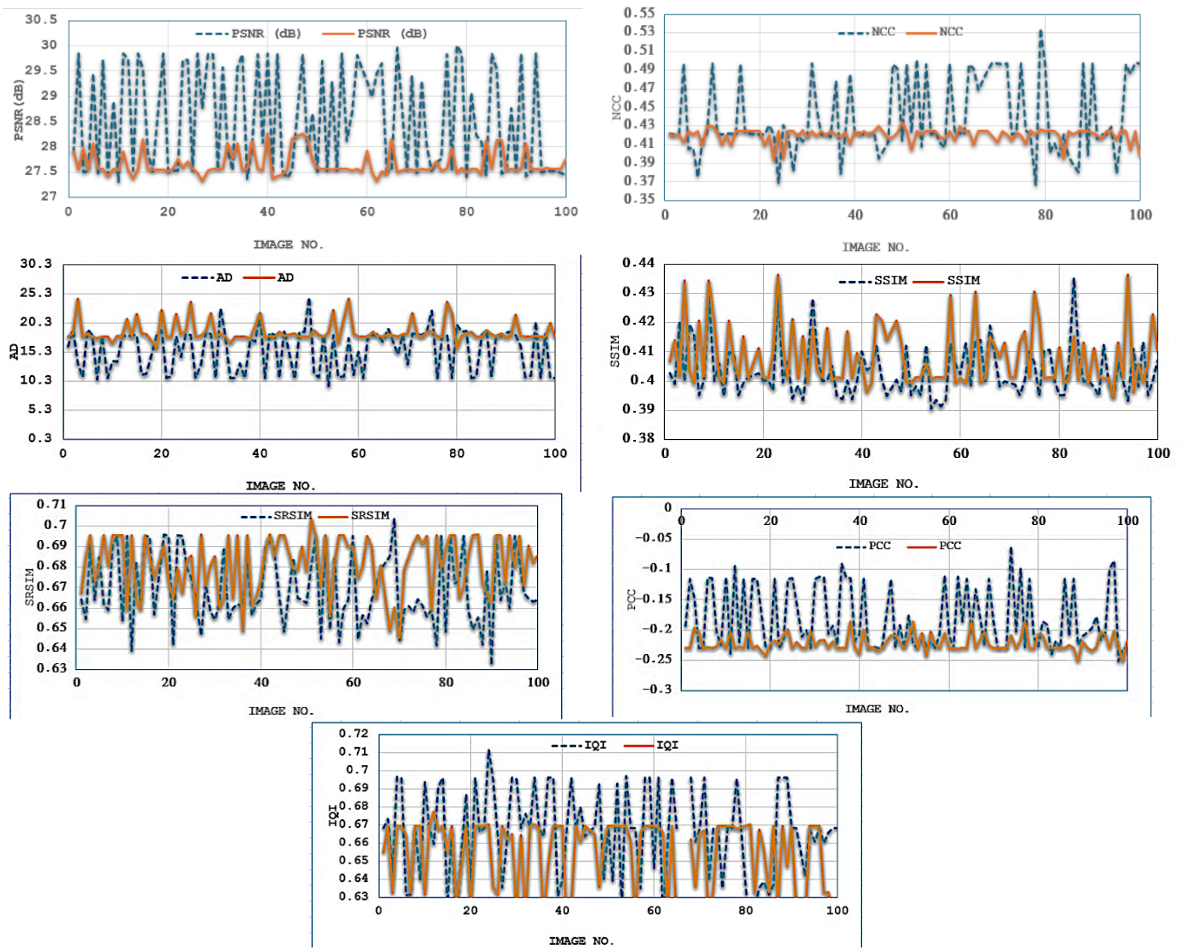

Figure 5: Acceptable results based on the image quality metrics.

5 DISCUSSION THE RESULTS

MATLAB (2021a) employ in the proposed method. A Core i5 processor and 8 gigabytes of RAM make up the system resources-processing and extraction for IQM to be done. A Windows application created in Python 3.9/64-bit with the Library programming language is the other operating system that has been offered and used in this paper to train the classifiers. DNN and RF were the classification approaches employed in training.

The image quality metrics used in the current study revealed significant differences in the diagnosing and differentiating breast diseases from healthy cases as described, Table 2.

Table 2: The performance metrics of the proposed methods using different classifiers.

Methods	DNN	RF
Acc	0.90	0.86
Prec	0.92	0.90
Recall	0.90	0.88
F1 measure	0.93	0.88

This study was conducted to identify breast tumors using computer-aided video (CAD) analysis of medical data from specialized surgical hospitals and laboratories. One hundred healthy and one hundred tumorous cases were collected for thermal imaging. The selection of appropriate features was considered more important than the selection of the classifier, and this was achieved using a number of characteristics.

The efficiency of the proposed method was evaluated by comparing it to previous studies. The proposed DNN method outperforms other related methods based on traditional machine learning approaches. It is an effective tool that can be used to assist clinicians in the rapid diagnosis and detection of breast tumors. Both the cost and speed of diagnosis will be improved.

In the final comparison of the outcomes obtained by the different prediction algorithms, DNN outperformed RF, which yielded the lowest result.

6 CONCLUSIONS

This study presents an AI-driven classification framework for distinguishing normal from abnormal breast thermal images using Deep Neural Networks (DNN) and Random Forest (RF) algorithms. Quantitative evaluation demonstrated that DNN outperformed RF, achieving a classification accuracy of 90.25% compared to 86.65%, respectively. The superior performance of DNN is attributed to its capability to automatically extract hierarchical thermal features from raw infrared data, whereas RF relies on manually defined texture metrics such as GLCM and asymmetry analysis. Image quality indices-including PSNR, SSIM, and NCC-confirmed the reliability of the preprocessing pipeline. From a biophysical perspective, abnormal thermal patterns reflect localized metabolic variations and vascular asymmetries associated with tumor-induced hyperthermia. The proposed methodology reduces diagnostic time and operator dependency, offering a non-invasive, radiation-free adjunct to conventional mammography. While thermal imaging alone is not a standalone screening tool, integration with DNN-based classification enhances early detection potential. Future work will extend this approach to multi-class thermal pathology diagnosis and validate across larger, multi-center datasets to improve generalizability and clinical applicability.

ACKNOWLEDGMENTS

This work was supported by Mustansiriyah University (www.uomustansiriyah.edu.iq).

REFERENCES

[1] M. A. Khan, S. Rubab, A. Kashif, M. I. Sharif, N. Muhammad, J. H. Shah, et al., "Lungs cancer classification from ct images: an integrated design of contrast based classical features fusion and selection," Pattern Recognition Letters, vol. 129, pp. 77-85, 2019.

[2] R. L. Sultan, "Diagnosis of Type II Diabetes Based on Feed Forward Neural Network Techniques," International Journal of Research in Pharmaceutical Sciences, vol. 11, no. 1, pp. 1109-1116, 2020.

[3] B. Hu, N. El Hajj, S. Sittler, N. Lammert, R. Barnes, and A. Meloni-Ehrig, "Gastric cancer: classification, histology and application of molecular pathology," Journal of Gastrointestinal Oncology, vol. 3, p. 251, 2012.

[4] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1790-1821, 2017.

[5] K. Fox, "Thermography and Breast Cancer Screening," Apr. 29, 2025.

[6] S. Rodriguez-Guerrero, H. Loaiza-Correa, A. D. Restrepo-Girón, L. A. Reyes, L. A. Olave, S. Diaz, and R. Pacheco, "Dataset of breast thermography images for the detection of benign and malignant masses," Data in Brief, vol. 54, 110503, 2024.

[7] A. Y. Vestergaard, J. Macaskill, T. R. Holt, and D. M. McKinlay, "The risk of melanoma in people with common melanocytic naevi: a systematic review," British Journal of Dermatology, vol. 157, no. 1, pp. 1-18, 2007.

[8] K. Zhang, L. Zhang, and H. Wang, "Local average pattern for texture classification," Pattern Recognition, vol. 43, no. 6, pp. 2005-2011, 2010.

[9] Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, and S. Zhao, "Skin cancer classification with deep learning: A systematic review," Frontiers in Oncology, vol. 12, 893972, 2022.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, 2017.

[11] P. Hermosilla, R. Soto, E. Vega, C. Suazo, and J. Ponce, "Skin cancer detection and classification using neural network algorithms: A systematic review," Diagnostics, vol. 14, no. 4, p. 454, 2024.

[12] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, and H. Kittler, "Human-computer collaboration for skin cancer recognition," Lancet Digital Health, vol. 2, no. 8, 2020.

[13] Z. M. Abood and N. S. Ali, "Thermal Images Modeling for Identify Abnormalities Regions in Breast," IJournals: International Journal of Social Relevance & Concern, vol. 4, no. 5, May 2016.

[14] A. H. Ornek and M. Ceylan, "Explainable Artificial Intelligence (XAI): Classification of Medical Thermal Images of Neonates Using Class Activation Maps," Traitement du Signal, vol. 38, no. 5, pp. 1271-1279, Oct. 2021.

[15] A. Z. Nowakowski and M. Kaczmarek, "Artificial Intelligence in IR Thermal Imaging and Sensing for Medical Applications," Sensors, vol. 25, no. 3, p. 891, 2025, [Online]. Available: https://doi.org/10.3390/s25030891.

[16] K. Kumar, M. Pradeepa, M. Mahdal, S. Verma, M. V. L. N. RajaRao, and J. V. N. Ramesh, "A deep learning approach for kidney disease recognition and prediction through image processing," Applied Sciences, vol. 13, no. 6, p. 3621, 2023.

[17] "Thermal imaging," [Online]. Available: https://visual-ic-uff-br.translate.goog/.

[18] "Breast Thermography," [Online]. Available: https://data.mendeley.com.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[20] T. G. Debelee, "Skin lesion classification and detection using machine learning techniques: A systematic review," Diagnostics, vol. 13, no. 19, p. 3147, 2023.

[21] B. A. Mohammed and Z. M. Abood, "Multibiometric Classification for People Based on Artificial Bee Colony Method and Decision Tree," AIP Conference Proceedings, vol. 2845, no. 1, 050009, 2023.

[22] A. Abedalla, M. A. Elsayed, and H. H. Ali, "Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review," Diagnostics, vol. 11, no. 8, 2021.

[23] R. E. Al-Bayati and Z. M. Abood, "Human gait features extraction based on angles and curves," AIP Conference Proceedings, vol. 2834, no. 1, 050010, 2023.

[24] S. Ray, "Disease Classification within Dermascopic Images Using Features Extracted by ResNet50 and Classification through Deep Forest," arXiv preprint, arXiv:1807.05711, 2018.

[25] B. Shabbir, M. Sharif, W. Nisar, M. Yasmin, and S. L. Fernandes, "Automatic cotton wool spots extraction in retinal images using texture segmentation and gabor wavelet," Journal of Integrated Design and Process Science, vol. 20, pp. 65-76, 2016.