Advances in Artificial Intelligence and Machine Learning; Research 5 (2) 3923-3953 Received 7-04-2025; Accepted 19-06-2025; Published 26-06-2025

Evaluating the Effect of Noisy Thermal Images On the Detection of Early Breast Cancer Using Deep Learning

Mohammed Abdulla Salim Al Husaini mohammed.h@aou.edu.om
Faculty of Computer Studies,
Arab Open University, Muscat 130, Oman

Mohamed Hadi Habaebi habaebi@iium.edu.my
IoT and Wireless Communication Protocols Laboratory,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia (IIUM),
Kuala Lumpur 53100, Malaysia.

Md Rafiqul Islam
IoT and Wireless Communication Protocols Laboratory,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia (IIUM),
Kuala Lumpur 53100, Malaysia.

Yousuf Nasser Said Al Husaini
Faculty of Computer Studies,
Arab Open University, Muscat 130, Oman

Corresponding Author: Mohammed Abdulla Salim Al Husaini

Abstract

Breast cancer remains a leading cause of mortality among women globally. Some techniques have been developed to enhance early detection, among which thermal imaging has emerged as a promising modality capable of identifying potential signs of breast cancer in its early stages. In addition, Thermal images provide valuable pixel-level information by capturing temperature variations between healthy and cancerous tissues. However, the susceptibility of these thermal images to noise poses a challenge to diagnostic accuracy in the early stages. This research aims to assess the influence of various types of noise on the performance of recently developed deep-learning models designed for early breast cancer detection. In addition, a comprehensive analysis was conducted using a substantial dataset to assess the impact of noise on the models' efficacy. It also encompasses different categories of noise, which are characterised by distinct mean and variance values ranging from 0.01 to 0.09. The findings reveal that introducing various types of noise, albeit within a small range of mean and variance values, adversely affects the performance of deep learning models. It shows that these filters play a pivotal role in enhancing classification accuracy. Moreover,

Citation: Mohammed Abdulla Salim Al Husaini, et al. Evaluating the Effect of Noisy Thermal Images On the Detection of Early Breast Cancer Using Deep Learning Advances in Artificial Intelligence and Machine Learning. 2025;5(2):222.

the results show that salt and pepper noise varied between 0.1 and 0.3, significantly impacting the accuracy of Inception MV4, reducing it from 100% to 51.58%, without adding filters in pre-processing. Additionally, introducing variance in multiplicative noise from 0.2 to 0.8 demonstrated an effect on classification accuracy only at noise levels of 0.7 (89%) and 0.8 (43%). Moreover, the results show that performance metrics for the proposed method were accuracy of 99.975%, sensitivity of 0.994, specificity of 1, precision of 1, NPV of 0.995, FNR of 0.006, LRN of 0.006, AUC of 0.997, EER of 0.003, and F1 score of 0.997, but FPR of 0. In conclusion, findings underscore the significance of refining noise mitigation strategies and pre-processing techniques to advance the reliability and accuracy of thermal imaging as a diagnostic tool in breast cancer detection in the early stages.

Keywords: Breast cancer, Thermal image, Gaussian noise, Deep learning.

1. INTRODUCTION

According to the World Health Organization, statistics indicate an increase in the incidence of breast cancer. The statistics mentioned in 2020 suggest that there are 2.088 million new cases of breast cancer, and the most crucial reason is late detection [1]. Machine learning and deep learning have emerged as transformative technologies in the medical sciences, revolutionising various aspects of healthcare, including disease diagnosis in early stages and treatment [2]. In addition, these advanced techniques have played a pivotal role in improving accuracy and efficiency in the realm of breast cancer detection [3–5]. Moreover, with vast amounts of accurate medical datasets available, machine learning algorithms can analyse intricate patterns and anomalies within mammograms, ultrasound images, and pathology slides, aiding in the early detection of breast cancer. Deep learning models have demonstrated remarkable capabilities in image recognition and classification tasks, enabling healthcare professionals to identify suspicious lesions with high precision. Moreover, these technologies have the potential to assist radiologists and pathologists by providing them with decision-support tools, thereby enhancing diagnostic accuracy and reducing the likelihood of human error [6]. Recently, thermal imaging technology has been utilised to detect breast cancer using deep learning. Therefore, it is a safe, harmless, and promising technique for early detection [7]. This technology has become a home diagnostic tool, enabling patients to track tumour status [8], periodically. Therefore, it is subject to noise that may sometimes not comply with thermal imaging procedures. Therefore, one of the factors affecting the classification of thermal images in deep learning is the location and size of the tumour in the breast and the presence of noise in thermal images. This study will explain the impact of Gaussian Noise on thermal images and datasets. Image noise, undesired information within an image, can manifest at various stages, like image capture, transmission, or processing. In addition, a comprehensive understanding of the noise's characteristics is essential to effectively eliminating noise from a noisy image. On the other hand, attempts at noise removal may lead to image blurring [9]. The study in [10], explores the diverse effects of noise types, such as Gaussian, Salt & Pepper, Speckle, and Poison, on thermal image features, specifically Histogram, Texture, and Transform-based features. Evaluating both pre- and post-noise applications, signal-to-noise ratio and mean absolute error serve as quality criteria, and results underscore the substantial influence of the mode feature on thermal images compared to regular images.

The fundamental idea is to enhance the conventional Laplacian of Gaussian filter by incorporating interval analysis to account for intensity uncertainties. Also, experimental evaluations were conducted on MIAS, a widely used breast cancer dataset, for medical image segmentation. Finally, system performance was assessed compared to Prewitt, LoG, and Canny filters using the Peak Signal-to-Noise Ratio (PSNR) [11]. On the other hand, images captured from a Mammogram might exhibit noise caused by variations in lighting and sensor inaccuracies. Cancer's noise can be effectively removed without compromising the image's boundaries and small details, and accurate diagnoses of breast cancer can be facilitated through imaging technology [12]. The presence of Salt and pepper Noise in thermal imaging may not be the most prevalent type of noise; it can occur under certain conditions. However, Salt and pepper noise can result from faulty sensor pixels, data transmission errors, or sudden environmental disturbances such as dead pixels, hot pixels, or random bit errors during image acquisition and transmission, which can introduce isolated bright or dark spots in thermal images [13]. Moreover, thermal cameras operating in harsh or unstable environments might experience impulsive noise due to abrupt temperature changes or electromagnetic interference [14]. By considering salt and pepper Noise in our evaluation, we aim to test the quality of our method against a range of possible noise artefacts that could affect thermal imaging systems, especially under specific conditions.

Breast cancer can be detected using various methods, including thermography, mammography, CT scans, MRI, ultrasound, and biopsy. However, mammography has limitations, as it exposes the body to ionising radiation and faces challenges in identifying cancer in dense breast tissues [15]. Furthermore, in some countries, mammography is solely used as a screening method. On the other hand, thermography offers a non-invasive and cost-effective alternative that relies on analysing temperature distribution on the breast surface. Its principles are rooted in the consistent emissivity of human skin temperature and detecting infrared radiation emitted by objects with temperatures above absolute zero, allowing for mapping the skin's temperature profile based on emitted infrared radiation [16].

Infrared imaging (thermography) and traditional mammography offer distinct approaches to breast cancer detection, each with unique advantages and limitations. Mammography, the gold standard since the 1960s, uses X-rays to identify structural abnormalities and has reduced mortality rates by 25% through early detection. However, it exposes patients to ionizing radiation, struggles with dense breast tissues, and may cause discomfort during compression. In contrast, thermography is non-invasive, radiation-free, and painless, capturing temperature variations on the breast surface to identify metabolic and vascular changes associated with tumors. Studies cited in the paper highlight its potential for early detection, even in cases where mammography failed, such as a 1996 comparison where thermal imaging detected a tumor missed by X-rays. Thermography is particularly advantageous for younger women or those with dense breasts. However, its diagnostic accuracy is highly dependent on strict imaging protocols (e.g., controlled room temperature, patient preparation) and is susceptible to noise artefacts, which can degrade performance. While some studies report high accuracy (80–100%), these results often derive from small datasets, raising concerns about overestimation. Additionally, thermography currently lacks the standardization and widespread clinical validation of mammography. Despite these challenges, thermography is a promising complementary tool, offering a safer alternative for routine monitoring and early screening. However, it is not yet positioned to replace mammography in definitive diagnostics [17].

1.1 Related Works

The study [18], referred to the use of the numerical simulation system for the location of tumours in the breast on different sites and the addition of thermal cooling. Gaussian Noises were added to thermal images. In addition, only four tumour sites were placed in the breast at different positions. The obtained results claim that the time-domain phase approach shows improved detection capabilities over those of raw heat. The study conducted by the researcher [19], proposed a new model of DCNN called SPARs. This model has been proposed for low-dimensional deep heat extraction. In addition, 208 clinical breast cancer thermal images were used with the DMR IR dataset. Moreover, Gaussian Noise was added at a rate of 3 to 20% on the thermal images, but after encoding, the results showed a slight decrease in the accuracy of the thermal images. SPAER demonstrated excellent robustness when tested for additive Gaussian Noise conditions (3-20% Noise), as assessed by its signal-to-noise ratio (SNR). The results indicate a high performance of SPAER in maintaining thermal heterogeneity, and it can be used as an in vivo non-invasive tool that aids CBE in the early detection of breast cancer.

In [20], Deep CNN and BIOS optimisation algorithms were used. The total thermal images used was 3895, of which 3098 were healthy breast images and 757 were cancerous breast images. Image processing with added noise was used. The results indicate an increase in the accuracy rate from 97.91 to 98.95%. The researcher [21], processed thermal images and removed noise using the soft wavelet threshold before colour segmentation. In addition, the proposed method was applied to 50 thermal images after applying pre-treatment. Clinical thermal images were obtained using a 640 x 480 pixels thermal camera. The results indicate that carefully designed colour maps visually improve the thermal image and improve pest detection and interpretation without using algorithms or deep learning. The study presented in [22], utilised Cascade Deep CNN to classify thermal images of breast cancer. The dataset consists of 900 thermal images. Gaussian Noise has been added to increase the dataset of thermal images. In addition, Deep CNN was trained on thermal images for automatic classification. CNN was trained by setting 50 epochs and a learning rate of 0.5e-3. The results indicated that the accuracy could reach 92%.

The researchers in [23], utilised the DCNN BreaCNet model, a modified version of the DCNN Shuffle Net model, which can extract around 6 million features. In addition, 1302 thermal images were used for training 90% and testing 10%, which were collected from the DMR IR dataset. Moreover, the Sobel kernel was used to filter the edges in the region of interest (ROI) and convert colour thermal images into grayscale. In addition, the modified model DCNN was tuned, learning 1e-3 with 75 epochs, and the SGDM optimisation method was used. Results indicated that the detection accuracy reached 100% compared to Mobilnet, which reached 98%. Furthermore, a Gaussian filter and a smartphone app with a size of 22MB have been added.

In [24], a new breast cancer detection method, DBC-4D U-Net-DITI, is introduced, utilising digital infrared thermal imaging. It employs Altered Phase Preserving Dynamic Range Compression (APPDRC) for pre-processing, optimises 4D U-Net weights using Glow-worm Swarm Optimization Algorithm (GSOA) for segmentation, reduces speckle noise, and employs a Binarized Spiking Neural Network (BSNN) for pathology stage classification. In conclusion, results indicate that in comparison to existing methods like DBC-CSSA-DITI, DBC-MPA-DITI, and DBC-CNN-DITI, the proposed approach demonstrates substantial improvements with a 39.01%, 28.34%, and 37.45% accuracy boost and a 17.12%, 24.12%, and 32.07% precision enhancement, respectively.

Thermal medical imaging, which captures infrared radiation emitted by the body, is susceptible to various noise types that degrade image quality. Below is a structured argument for the existence of Speckle, Gaussian, Salt & Pepper, and Poisson noise in these images, supported by scientific references.

Speckle noise, typically associated with coherent imaging (e.g., ultrasound), can manifest in thermal imaging due to sensor non-uniformities or scattering effects. While thermal imaging is passive and incoherent, variations in detector responsivity across microbolometer arrays create multiplicative patterns resembling Speckle noise. Surface roughness in imaged tissues or sensor coatings may also scatter infrared waves, introducing granular interference [25, 26]. In sensor circuits, gaussian (additive white) noise arises from electronic thermal noise (Johnson-Nyquist noise). This noise, caused by random electron motion in resistive components (e.g., microbolometer readout circuits), follows a normal distribution and is temperature dependent [27]. Salt & Pepper (impulsive) noise manifests as random bright/dark pixels due to sensor defects (dead/broken pixels) or transmission errors. Faulty microbolometer elements or electromagnetic interference during data transfer can cause transient pixel saturation [28, 29].

Poisson (photon) noise stems from quantum fluctuations in detected infrared photons. While thermal detectors (microbolometers) measure temperature-dependent resistance, the incident photon flux follows Poisson statistics, where variance equals the signal mean. This noise is prominent in low-flux scenarios (e.g., imaging cool body regions) [30, 31].

In summary, thermal medical images are inherently affected by multiple noise types, due to sensor physics, electronic limitations, and photon statistics. Speckle noise arises from detector non-uniformities, Gaussian noise from electronic thermal agitation, Salt & Pepper noise from sensor/transmission faults, and Poisson noise from quantum fluctuations. Addressing these noises is critical for enhancing diagnostic accuracy in applications like inflammation detection or cancer screening.

It is clear from the previous studies and TABLE 1, that Gaussian Noise was used. However, these studies lack clarity of noise data. Therefore, the proposed approach introduces three different types of noise with varying values to a dataset consisting of thermal images. In addition, the effect of an added set of noise values on the classification capabilities of DCNN Inception MV4 is studied. In conclusion, filters were added to mitigate the impact of noise from thermal images before classification.

1.2 Motivation

Breast cancer is a significant health concern worldwide, and early detection contributes to improving patient outcomes. Thermal imaging, as a non-invasive method, shows promise in detecting breast cancer in early stages. However, thermal images are susceptible to different types of noise, such as Gaussian Noise, speckle noise, salt and pepper noise, and Poisson noise, affecting image quality. Therefore, understanding the effects of different types of noises on thermal images of breast cancer is essential for developing robust and reliable deep-learning algorithms for accurate detection and diagnosis. The novelty of this work lies in evaluating the performance of the DCNN MV4 deep learning model on noisy thermal images.

Table 1: Comparison between different Noise type and Deep learning models for early breast cancer detection.

Source	Noise type	Deep learning model	Accuracy
[8]	white Gaussian noise	NA	NA
[9]	Gaussian noise (3–20% noise)	sparse deep convolutional autoencoder (SPAER) model	78.2%
[10]	salt and pepper noise	CNN	98.95%
[11]	Poisson noise	NA	NA
[12]	Gaussian noise	cascaded CNN architecture	92%
[13]	Normal noise	BreaCNet	100%
[14]	speckle noise	Unet	improvements with a 39.01%

NA = Not Available

1.3 Contribution

The work in this paper on the evaluation of Gaussian noise effects in thermal images of breast cancer and its impact on deep learning methods has manifold contributions:

(i) The study thoroughly evaluates the effects of different types of noises on thermal images, specifically related to breast cancer. Studying the effect of noise on image quality, essential features, and overall diagnostic performance enhances our understanding of the challenges associated with noisy thermal images.

(ii) The research examines the performance of deep learning algorithms, such as convolutional neural networks (CNNs), in Gaussian Noise. Quantifying the degradation in accuracy, sensitivity, and specificity caused by noise provides insights into the limitations of existing deep-learning models and highlights the need for noise-aware techniques.

(iii) Based on the evaluation results, the study investigates and proposes effective strategies to mitigate the adverse effects of different types of noise on thermal images. This could include pre-processing techniques, noise reduction algorithms, or modified deep learning architectures that are robust to noise, ultimately improving the accuracy and reliability of breast cancer detection systems.

(iv) The findings and recommendations from this research have practical implications for medical professionals, researchers, and developers working on thermal imaging-based breast cancer diagnosis. The study can guide the development of noise-robust deep learning models, contribute to the design of optimised imaging protocols, and assist in implementing quality control measures for thermal imaging devices.

By addressing the motivation and making these significant contributions, the evaluation of Gaussian noise effects in thermal images of breast cancer and deep learning advances our understanding of the challenges associated with noisy thermal images. It helps pave the way for improved early detection and diagnosis of breast cancer.

1.4 Objective

The objective of this study is to evaluate the impact of various types of noise—including Gaussian, salt and pepper, speckle, and Poisson noise—on the performance of the deep learning model Inception MV4 in detecting early-stage breast cancer using thermal imaging. By systematically introducing noise with different mean and variance values, the research aims to quantify how these disturbances degrade classification accuracy and diagnostic reliability. Additionally, the study seeks to propose effective noise mitigation strategies, such as pre-processing filters and denoising algorithms, to enhance the robustness of deep learning models and improve the accuracy of thermal imaging as a non-invasive diagnostic tool for early breast cancer detection.

2. MATERIALS AND METHODS

This section will introduce the deep learning framework used in this paper. Namely, inception v3, inception V4, and modified inception V4 models [32], are introduced below.

2.1 Inception V3

Inception-v3 is a CNN architecture created in 2015 by Google researchers. It is part of the Inception family of CNN models and is designed for image recognition and classification tasks. Inception v3 is characterised by its deep architecture, consisting of 48 layers in total. It uses a combination of different sizes of filters in convolutional layers, max pooling, average pooling layers, and fully connected layers. The architecture is based on "inception modules," which are building blocks that allow for efficient multi-scale feature extraction. The main innovation of Inception-v3 lies in its inception modules. These modules use various filter sizes (1x1, 3x3, 5x5) to capture information at different spatial scales and process them in parallel. This helps the network capture both fine-grained details and high-level features. To reduce the computational complexity of the network, Inception v3 uses 1x1 convolutions as dimensionality reduction layers. These layers minimise the number of input channels before applying larger convolutional filters, thereby improving efficiency without significant loss of information. Inception v3 includes auxiliary classifiers at intermediate layers. These classifiers help combat the vanishing gradient problem during training and provide regularisation. They also serve as additional sources of gradients during backpropagation, aiding the overall training process. Inception-v3 is typically pre-trained on large-scale datasets like ImageNet, which contains millions of labelled images from various categories. After pre-training, the model can be fine-tuned on smaller datasets for specific image recognition tasks, adapting its learned features to the new dataset. Inception-v3 achieves impressive performance on image classification benchmarks. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 achieved a top 5 error rate of 3.46%, surpassing its predecessor models and demonstrating state-of-the-art accuracy. Inception v3 and its subsequent versions have been widely adopted and used as a backbone architecture in many computer vision applications, including object recognition, image segmentation, and transfer learning tasks. Its design principles have also influenced the development of other CNN architectures, such as Inception-V4 and Inception-ResNet [33].

2.2 Inception V4

Inception V4 is an advanced architecture that improves upon previous versions of the Inception family. It simplifies the overall structure, introduces a stem layer, and incorporates more inception modules than Inception-v3. The architecture of Inception V4 is characterised by a more uniform and simplified design. The network comprises a comprehensive planner and stem configuration, four inception A layers, seven inception B layers, and three inception C layers [34].

2.3 Inception MV4

An examination was conducted to compare "Inception B" in Inception V4 to the updated version in MV4. The modifications were limited to "inception B", as shown in FIGURE 1. First, a convolutional layer was added beneath the average pooling layer. The number of filters was increased from 128 to 256 to maintain the number of features extracted. Then, two parallel convolutional layers were added under the layer with 192 filters, with the same size and number of filters. After that, the remaining layers in Inception B were removed to keep the number of extracted features. Finally, Inception B has seven groups, each with a set of layers. All seven groups were changed with the same adjustments mentioned before. For further details on the framework, readers are advised to consult [32].

2.4 DMR IR Dataset for Mastology Research with Infrared Image

The DMR IR (Dataset of Mastologic Research with Infrared Images) is a specialized dataset designed for breast cancer detection using infrared thermography. It comprises 1,800 thermal images, categorized into 800 healthy and 1,000 cancerous cases, to support machine learning applications. The patient must abstain from hot liquid activity and apply any creams to their breasts and underarms at least 2 hours before the exam. Moreover, the exam room should be kept at a temperature between $20 - 22°C$. Also, Patient Preparation: In the exam room, the patient should remove any jewellery or accessories that could affect the thermal image. Additionally, the camera should be positioned 1 meter from the patient. The acquisition should be dynamic, with the patient facing the camera. IR images are taken using a FLIR SC620 with a sensitivity of less than $0.04°C$ and a temperature range of $-40°C$ to $500°C$. It can be accessed through an online interface [35], that allows for easy management and retrieval of information from breast exams and patient clinical data.

2.5 Speckle Noise Model

A speckle pattern that appears in polarised monochromatic light can be viewed as being caused by a classical random walk in the complex plane [36]. Speckle noise is a type of noise that appears in images as a rough and grainy pattern resembling salt and pepper noise. It has the most significant impact on thermal image features, while Salt and pepper noise has a lower effect [10]. This noise is referred to as a multiplicative noise with a granular form. The presence of speckle noise in images is unwanted as it degrades the thermal image quality by impacting the edges and local details between different organs, which are crucial for diagnostic purposes [37]. Speckle noise

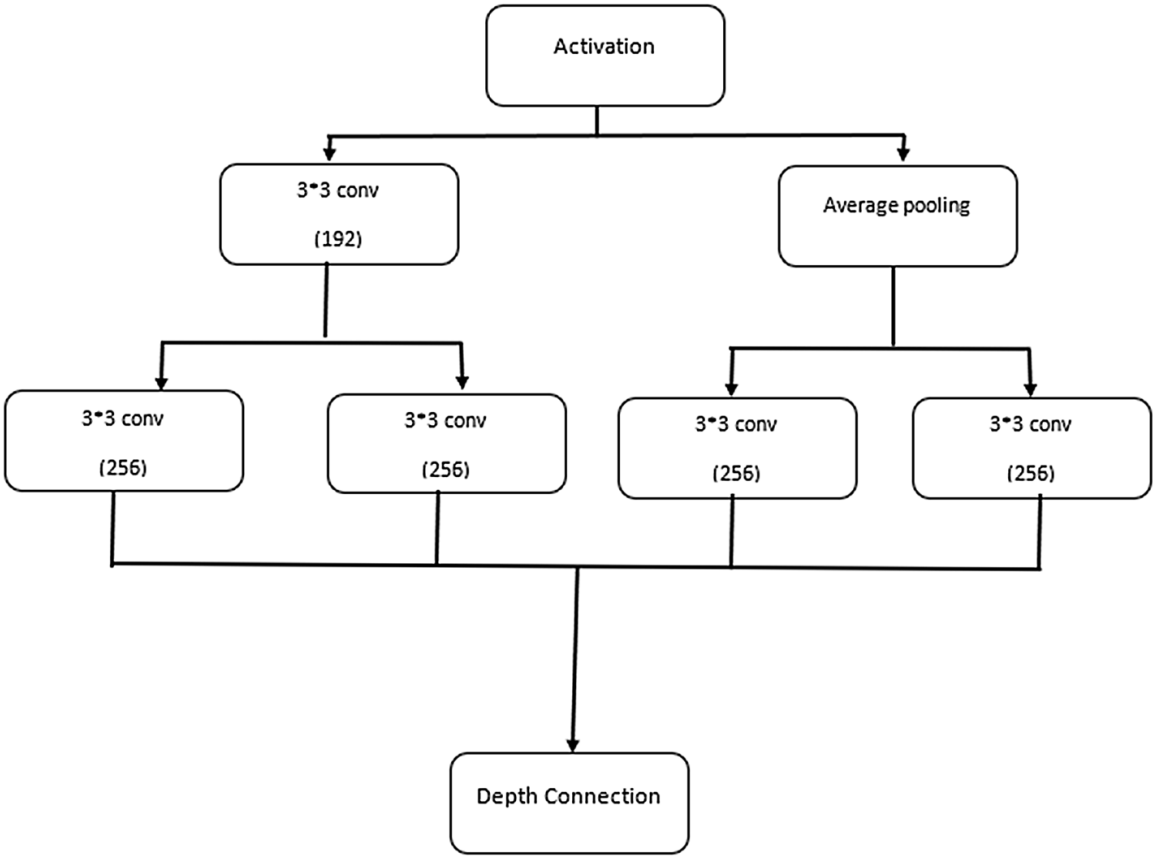

Figure 1: Inception MV4

in the medical literature is known as "texture" and might hold valuable diagnostic data. The ideal level of speckle smoothing is primarily influenced by the expert's expertise and the intended use. In automatic segmentation, preserving the crispness of boundaries between several image areas is commonly selected while reducing the speckled texture [38].

2.6 Gaussian Noise Model

Gaussian Noise, also known as additive white Gaussian Noise (AWGN), is a common type of random noise in many imaging and signal-processing applications. It is characterised by random variations in pixel intensities that follow a Gaussian distribution. It is often introduced during image acquisition, transmission, or processing due to electronic noise, quantisation errors, or sensor imperfections. The noise appears as a random fluctuation around the actual pixel value, resulting in a smooth distribution of noise values across the image. Moreover, the properties of Gaussian Noise make it a vital noise model in many theoretical and practical applications. In addition, the Gaussian distribution is symmetric and bell-shaped, with most noise values concentrated around the mean value and fewer values occurring at the extremes. However, the noise values are statistically independent at each pixel, and the mean and standard deviation parameters characterise their distri-

bution. The choice of denoising technique depends on the noise characteristics, the desired level of noise reduction, and the trade-off between noise removal and preservation of image details [39].

2.7 Salt & Pepper Noise

Salt and pepper noise, also known as impulse noise, is a common image noise that can affect the quality and clarity of digital thermal images. It is characterised by randomly scattered white and black pixels resembling grains of salt and pepper. It can occur due to various factors, such as transmission errors, faulty sensors, or electronic interference during the image acquisition or processing stages. It typically manifests as isolated bright white pixels (salt) and dark black pixels (pepper) randomly scattered throughout the image. Salt and pepper noise can significantly affect image analysis and processing tasks, as it introduces unwanted artefacts and distorts the image's overall appearance. Therefore, it is crucial to employ denoising techniques to mitigate the effects of this noise and restore image quality. When selecting a denoising method, it is important to consider the trade-off between noise removal and preservation of important image features. Different methods may yield varying results depending on the noise characteristics and the desired image quality [40].

2.8 Poisson Noise

Poisson noise is a random variation commonly occurring in images captured under low-light conditions or when using photon-counting sensors, like medical imaging devices. Poisson noise is caused by the random nature of photon arrival at the sensor or detector. In addition, it manifests as a variation in the number of photons detected at each pixel, resulting in intensity fluctuations in the image. The Poisson distribution models the probability of observing a certain number of events (photons) in each interval of time or space. In digital images, Poisson noise appears as random variations in pixel intensities, with brighter areas having a higher probability of more photons and darker areas having a lower probability. This means that the variance of Poisson noise is equal to the mean. This noise can degrade image quality by reducing contrast, introducing unwanted variations, and affecting image analysis and processing tasks. Moreover, due to its statistical properties, it is inherently different from other types of noise, such as Gaussian Noise. Therefore, specialised denoising techniques are required to address this specific type of noise [41], effectively.

2.9 Preprocessing

Pre-processing is a crucial step in breast cancer detection using thermal images to enhance the performance of DCNNs. This process involves resizing the thermal images to a suitable size and automatically removing unwanted areas, such as the armpits, neck, and abdomen, to focus on the region of interest. Additionally, data augmentation techniques are applied, including flipping the images to the left and right, flipping vertically, and rotating 30 degrees to the left and right to increase the diversity of the training data and improve the model's generalisation. The dataset is divided into training 70% and testing 30% subsets to ensure robust model development and evaluation. DCNN models are trained using optimised hyperparameters, such as a learning rate of 1e-4, batch size adjustments, and Stochastic Gradient Descent with Momentum (SGDM) as the optimisation

method to improve accuracy. These steps ensure that the model effectively extracts critical features from thermal images, leading to reliable and accurate detection of breast cancer in the early stages.

2.10 Experiment Setup

The experiments were conducted to evaluate the effect of different types of noise applied to the thermal images' dataset on the performance of the MV4 model in detecting breast cancer at an early stage. The results obtained were contrasted against similar results in the open literature. In addition, the experiments were divided into four scenarios. The experiments were conducted to evaluate the effect of different types of noise applied to the thermal images' dataset on the performance of the MV4 model in detecting breast cancer at an early stage. The results obtained were contrasted against similar results in the open literature. In addition, the experiments were divided into sixteen scenarios. The first scenario used the original clean dataset of thermal images from the DMR-IR repository as a benchmark to verify the performance of the modified inception MV4 model against other deep learning models. The second scenario used the Gaussian Noise-infected version of the thermal image dataset for experimentation with the MV4 model. The third scenario combines the clean dataset and the Gaussian Noise dataset. In the fourth scenario, a healthy thermal image was utilised as a baseline, to which various types of noise were systematically added at different intensity levels. This approach aimed to evaluate the impact of noise on the classification performance of the Inception MV4 model and to identify the threshold at which noise begins to affect classification accuracy significantly.

As for the devices used in the experiments, a computer with high specifications, 32 GB RAM, Core i7, and a storage capacity of 1 TB. In addition, MATLAB version 2020a software and graphics processor unit. Moreover, thermal images from the DMR IR dataset were utilised, with 1800 thermal images categorised as healthy (800 thermal images) and unhealthy (1000 thermal images). The deep CNN inception MV4 model was used because it has high accuracy compared to other neural networks, as shown in FIGURE 1 [32].

The learning rates (1×10^3, 1×10^4, 5×10^4) were selected through empirical tuning, guided by established practices in deep learning literature and iterative validation across optimization methods (SGDM, ADAM, RMSPROP). For instance, 1×10^4—a widely adopted default for CNNs—proved optimal for SGDM in Inception V4/MV4, achieving 100% accuracy while balancing convergence and stability. The batch size 10 was chosen to align with hardware constraints (32 GB RAM, 6 GB GPU) while maintaining gradient stability and computational efficiency. The epoch range (3–30) was determined empirically by monitoring validation performance, with training halted at plateaus (e.g., Inception V3 required 20–30 epochs to match V4's accuracy, likely due to architectural depth differences). Data augmentation parameters (30-pixel translation, 10% scaling) were derived from protocols in thermal imaging studies to preserve diagnostic features while ensuring invariance to minor variations [32]. In our study, we utilized a pre-trained deep learning-based denoising model known as DnCNN (Denoising Convolutional Neural Network), which is specifically designed to remove Gaussian noise from images while preserving essential structural information. The denoising process was applied separately to each RGB channel of the image. After noise removal, the channels were recombined to reconstruct the full-color image. This denoised image was then used as input for the classification model. The application of this method significantly improved the quality of the input data and contributed to more accurate classification results.

3. RESULTS AND DISCUSSION

Four distinct scenarios were studied. In the first benchmark scenario, Inception V3, Inception V4, and modified Inception MV4 DCNN models were trained using the same settings on the original dataset. Inception v3, Inception V4, and Inception MV4 utilise the same data augmentation techniques, including random flipping along the vertical axis, translating images up to 30 pixels, and scaling them horizontally and vertically by up to 10%. Also, the models were trained using the DMR-IR dataset, split into 70% for training and 30% for testing, with a learning rate of 1e-4 and a consistent configuration of freezing the first 10 convolutional layers.

3.1 Clean Dataset

The Inception MV4 model achieved the highest average accuracy (99.975%) and the lowest error margin (±0.18%), outperforming Inception V4 (99.971% accuracy, ±0.27% error) and Inception V3 (98.104% accuracy, ±1.52% error). Regarding training times, Inception MV4 had a moderate time of 7.704 minutes, while Inception V4 took the longest at 9.554 minutes. Inception V3, with fewer parameters, was trained in just 6.376 minutes but exhibited the highest error margin. These results highlight Inception MV4 as the most efficient model, effectively balancing performance, accuracy, and training time [32].

According to these results, we selected the DCNN MV4 model with the same settings, including a learning rate of 1e-4, 70% training and 30% testing dataset split, and consistent data augmentation techniques, as it demonstrated the highest accuracy (99.975%), the lowest error margin (±0.18%), and a balanced training time of 7.704 minutes, making it the most efficient and reliable choice for our application. As the Inception MV4 model was verified on the primary dataset, consisting of 1800 thermal images, the results indicate a detection accuracy of 99.975%, as shown in TABLE 2.

FIGURE 2, displays confusion matrices for Inception V3, Inception V4, and Inception MV4 models, highlighting their classification performance. Inception V3 achieves the highest True Positives 255 but has more False Positives 3 and False Negatives 15, indicating a trade-off between sensitivity and precision. On the other hand, Inception V4 and MV4 perform identically, with perfect precision and specificity. There are no false positives, but a single false negative 1. While V4 and MV4 minimise misclassifications, V3's higher recall makes it suitable for cases prioritising sensitivity. The matrices showcase the models' strengths and weaknesses in balancing accuracy and error rates.

To keep the focus of the paper on the noise effect, from now on, only the inception MV4 model will be considered in the rest of the studied scenarios. The optimised hyperparameter settings used for the MV4 model throughout the other scenarios are a learning rate 0.0001, the SGDM optimisation method, and a training of 10 epochs.

3.2 Gaussian Dataset

TABLE 3, indicates the detection accuracy of the inception MV4 model during 10 epochs under Gaussian noise conditions with variances of 0.05 and 0.02. In this dataset, Gaussian noise was applied to both classes healthy and cancerous images to evaluate the model's robustness across

Table 2: Performance Measures For inception MV4 Deep Learning Models.

Deep learning Models	**Inception v3**	**Inception V4**	**Inception MV4**
True Positive (TP)	170	171	171
True Negative (TN)	198	203	203
False Positive (FP)	2	0	0
False Negative (FN)	5	1	1
Accuracy	98.1	99.971	99.975
Sensitivity	0.971	0.994	0.994
Specificity	0.99	1.000	1.000
Precision	0.988	1.000	1.000
Negative Predictive Value (NPV)	0.976	0.995	0.995
False-Positive Rate (FPR)	0.01	0.000	0.000
False-Negative Rate (FNR)	0.029	0.006	0.006
Likelihood Ratio Positive (LRP)	97.1	NA	NA
Likelihood Ratio Negative (LRN)	0.029	0.006	0.006
Area Under Curve (AUC)	0.9805	0.997	0.997
Equal Error Rate (EER)	0.019	0.003	0.003
Harmonic mean of precision and recall (F1)	0.979	0.997	0.997

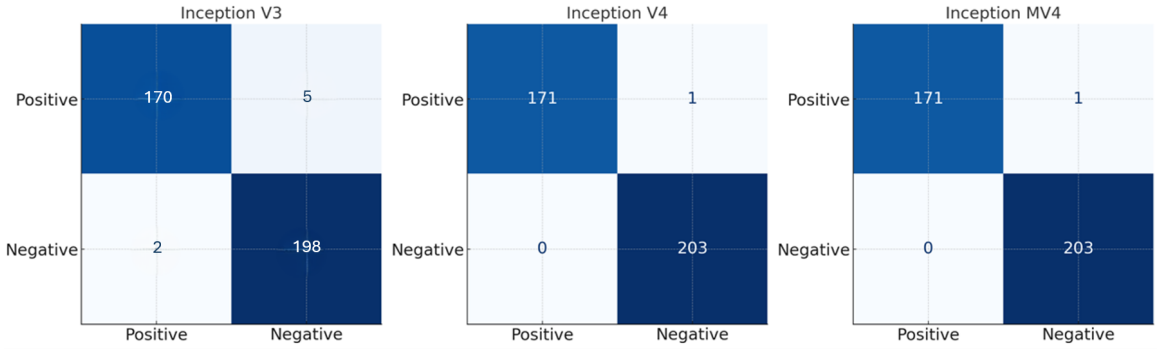

Figure 2: Confusion Matrices for Inception V3, V4, and MV4 Models

all image types. Initially, the accuracy starts at 96.06% for the first two epochs, then improves significantly to 96.96% in the third epoch. By the fourth epoch, the system achieves near-perfect accuracy at 99.82% and reaches 100% in the fifth and ninth epochs. Accuracy remains consistently high for the rest of the epochs, fluctuating slightly between 99.64% and 100%. Finally, the DCNN proposed model demonstrates strong robustness to Gaussian Noise, with an average accuracy of 98.8%, highlighting its reliable performance over time.

3.3 Combining Dataset

The DMR IR + Noise section evaluates the model's robustness under noisy conditions by using Gaussian noise to the DMR IR dataset. Moreover, only 50% of the original dataset where infected

Table 3: Average detection accuracy of different datasets.

Epoch	Gaussian Noise 0.05&0.02	DMR IR	DMR IR + NOISE
1	96.06	100	90.01
2	96.06	100	92.17
3	96.96	100	96.26
4	99.82	99.82	94.84
5	100	100	91.99
6	99.82	99.92	98.58
7	99.64	100	97.69
8	99.82	100	98.22
9	100	100	95.02
10	99.82	100	98.22

with Gaussian noise with a mean value of 0.05 and variance of 0.02. TABLE 3, shows the detection accuracy starts at 90.01% in epoch one and improves steadily, peaking at 98.58% in epoch 6, showcasing the model's ability to adapt and learn to handle noise. While the performance is slightly lower than the clean DMR IR dataset (100% accuracy), the results highlight the model's resilience in maintaining high accuracy levels despite variability. These fluctuations in accuracy across epochs suggest areas where noise impacts learning, providing opportunities for further optimization. FIGURE 3 illustrates the impact of Gaussian noise on detection accuracy using Inception MV4 for thermal images, where the noise mean ranges from 0.01 to 0.09 and variance from 0.01 to 0.09. The results show that increasing the noise variance significantly affects classification accuracy, leading to a higher likelihood of misclassifying. The model performs relatively well for lower noise levels (mean = 0.01, variance = 0.01–0.02) with stable detection accuracy. However, as the noise variance increases, the system becomes more prone to incorrect classifications, particularly for healthy images, which are mistakenly identified as cancer. Finally, noise can distort important features, leading to misclassifications or false detections.

3.4 Comparison Between Three Datasets

TABLE 3, presents the average detection accuracy across different epochs for three dataset configurations: Gaussian Noise (0.05 & 0.02), DMR IR, and DMR IR+NOISE. The Gaussian Noise dataset starts with lower accuracy at 96.06% for the first two epochs and gradually improves to 99.82% and 100% in later epochs. The DMR IR dataset maintains perfect detection accuracy (100%) across all ten epochs, demonstrating its robustness and reliability. The DMR IR+NOISE dataset also achieves near-perfect accuracy, beginning at 99.91% in the first two epochs and consistently reaching 100% from epoch three onwards. The average detection accuracy for these datasets reflects their relative performance: 98.8% for Gaussian Noise, 99.974% for DMR IR, and 98.22% for DMR IR+NOISE. When 50 % case the size of the dataset used for trained was not large enough for the system to produce high accuracy where as case fully noise was sufficient system to overcome to ignore the noise. Nevertheless, when the clean dataset the accuracy was much superior to other two cases This

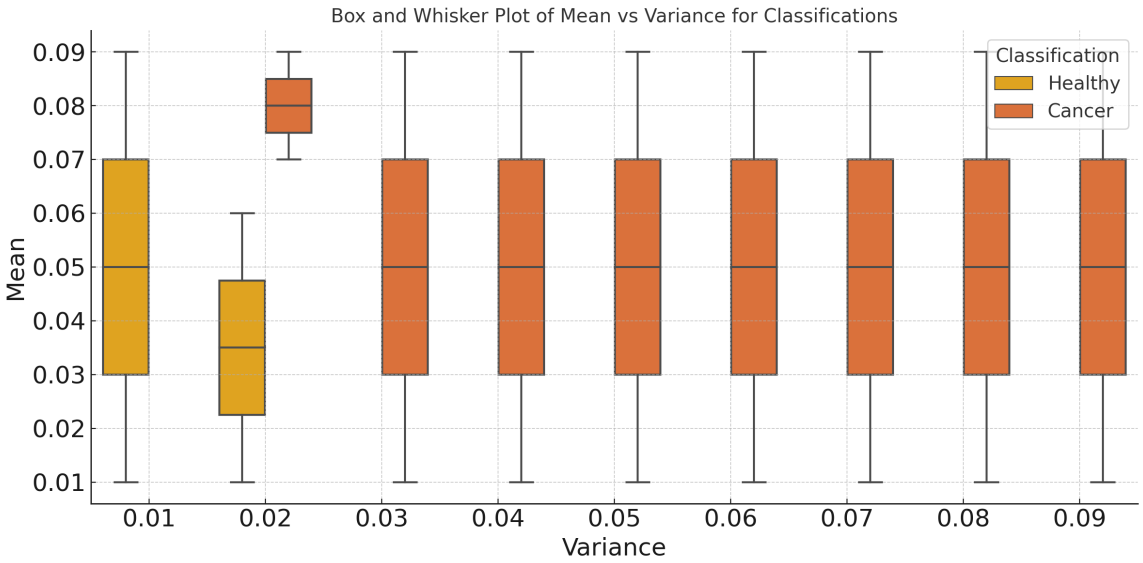

Figure 3: Detection accuracy for Thermal image by using Gaussian noise dataset only.

indicates that while Gaussian Noise shows slight variability, the DMR IR and DMR IR+NOISE configurations perform almost identically, with exceptional stability and reliability.

TABLE 4, evaluates various performance metrics for different datasets. The Gaussian Noise dataset achieves an accuracy of 99.82%, sensitivity of 1.000, specificity of 0.997, precision (P) of 0.996, and a negative predictive value (NPV) of 1.000. It also records false positive rate (FPR) and false negative rate (FNR) values of 0.003 and 0.000, respectively, with an AUC (area under the curve) of 0.998, equal error rate (EER) of 0.000, and an F1-score of 0.002. The IR+NOISE dataset records an accuracy of 98.2%, a sensitivity of 1.000, a specificity of 0.964, a precision of 0.965, and an NPV of 1.000. Its FPR and FNR are 0.036 and 0.000, respectively, achieving an AUC of 0.982 and an F1-score of 0.018. The DMR IR dataset achieves perfect results across all metrics, including accuracy, sensitivity, specificity, precision, NPV, and F1-score of 1.000, zero FPR, FNR, and EER values, and an AUC of 1.000. These results underscore the exceptional robustness of DMR IR compared to the Gaussian Noise and IR+NOISE datasets.

Table 4: Evaluation parameters for different type of dataset

Type of dataset	Accu	Sens	Spec	P	NPV	FPR	FNR	LRN	AUC	EER	F1
Gaussian Noise 0.05&0.02	99.82	1.000	0.997	0.996	1.000	0.003	0.000	291.000	0.000	0.998	0.002
DMR IR+NOISE	98.2	1.000	0.964	0.965	1.000	0.036	0.000	27.636	0.000	0.982	0.018
DMR IR	100	1.000	1.000	1.000	1.000	0.000	0.000	0.000	1.000	0.000	1.000

On the other hand, TABLE 5, shows that a statistical analysis using the T-test revealed significant differences in detection accuracy between Gaussian noise and both DMR IR ($p = 0.0418$) and DMR IR with Noise ($p = 0.0447$). Also, this indicates that noise significantly impacts performance compared to clean data. However, there was no significant difference between DMR IR and DMR

IR with Noise (p = 0.5109), suggesting that adding noise to DMR IR does not substantially affect performance.

Table 5: Pairwise T-Test Results for Detection Accuracy Across Different Datasets and Noise Conditions

Comparison	t-statistic	p-value
Gaussian Noise vs DMR IR	−2.191682	0.041795
Gaussian Noise vs DMR IR + Noise	−2.158039	0.044681
DMR IR vs DMR IR + Noise	0.670820	0.510852

3.5 Effect of Different Types of Noise on the Detection Accuracy of MV4 Model.

This section delves into the challenges of image noise in medical imaging analysis. Now, the focus shifts to speculative noise and salt-and-pepper disturbances, highlighting specific examples where such noise can influence classification outcomes.

3.5.1 Thermal image with speckle noise in different values

TABLE 6 presents the average detection accuracy for various values of speckle noise variance, with the mean value fixed at 0. The variance ranges from 0.02 to 0.08, and the table displays accuracy for five different settings. At lower variance levels of 0.02, 0.03, and 0.04, the accuracy across all five measurements remains perfect at 100%. However, as the variance increases, a slight decrease in accuracy is observed, dropping to 99.9958% at a variance of 0.05 and 99.9026% at a variance of 0.06. This decline becomes more pronounced at higher variance levels, with the accuracy dropping significantly to 89.6558% at 0.07 and further plummeting to 43.86% at 0.08. The individual accuracy settings highlight this trend, particularly accuracy 1, which decreases drastically to 0.6% at a variance of 0.08. These results indicate that while detection accuracy remains robust at lower variance levels, it is highly susceptible to degradation at higher variance levels, emphasising the importance of minimising variance in speckle noise to maintain performance. As the variance increases, thermal images in the classification begin to flip, being misclassified as cancer, with this shift occurring notably at a variance value of 0.08.

TABLE 7 illustrates the average detection accuracy for different mean values of speckle noise, with the variance fixed at 0.02. The mean varies from 0.02 to 0.08, and the results show consistently perfect detection accuracy at 100% for lower mean values of 0.02, 0.03, and 0.04. However, as the mean increases to 0.05, a slight decline in average accuracy is observed at 99.998%. Further increases in mean result in a noticeable drop in accuracy, with the average accuracy declining to 99.7518% at 0.06 and 98.94% at 0.07. The most dramatic decline occurs at a mean of 0.08, where the average accuracy plummets to 62.114%. Among the individual settings, accuracy two and three experience the most significant drops, reaching 44.5% and 3.3%, respectively, at a mean of 0.08. These findings highlight the sensitivity of detection accuracy to increases in the mean value of speckle noise, underscoring the need for effective noise reduction strategies to mitigate its effects and maintain high detection performance. As the mean increases, thermal images in the

Table 6: Average detection accuracy of different value in Variance of speckle noise

Setting (Mean = 0, variance is varied) (Detection Tipping Point @ variance = 0.08)							
Accuracy in	0.02	0.03	0.04	0.05	0.06	0.07	0.08
1^{st} detection attempt	100	100	100	100	99.996	99.8	0.6
2^{nd} detection attempt	100	100	100	99.98	100	99.11	94
3^{rd} detection attempt	100	100	100	99.999	99.999	99.999	30
4^{th} detection attempt	100	100	100	100	99.998	49.6	8
5^{th} detection attempt	100	100	100	100	99.52	99.77	86.7
Aaverage accuracy	100	100	100	99.9958	99.9026	89.6558	43.86

classification begin to flip to being misclassified, with this shift occurring notably at a mean value of 0.08.

Table 7: Average detection accuracy of different values in Mean of speckle noise

Setting (variance =0.02, Mean is varied) (Detection Tipping Point @ Mean = 0.08							
Accuracy in	0.02	0.03	0.04	0.05	0.06	0.07	0.08
1^{st} detection attempt	100	100	100	100	99.998	95.4	97.77
2^{nd} detection attempt	100	100	100	100	99.994	99.55	44.5
3^{rd} detection attempt	100	100	100	100	99.999	99.84	3.3
4^{th} detection attempt	100	100	100	99.99	98.8	99.99	69
5^{th} detection attempt	100	100	100	100	99.968	99.92	96
Average accuracy	100	100	100	99.998	99.7518	98.94	62.114

Table 8: Average detection accuracy of different values in Mean of Gaussian noise

Setting (variance =0.02, Mean is varied) (No Tipping Point)							
Accuracy in	0.02	0.03	0.04	0.05	0.06	0.07	0.08
1^{st} detection attempt	100	100	100	100	100	100	100
2^{nd} detection attempt	100	100	100	100	100	100	100
3^{rd} detection attempt	100	100	100	100	100	100	100
4^{th} detection attempt	100	100	100	99.99	100	100	100
5^{th} detection attempt	100	100	100	100	100	100	100
Average accuracy	100	100	100	99.998	100	100	100

A thermal images in FIGURE 4, represents an experiment involving medical thermal imaging, where the noise was artificially introduced and then filtered to assess its effect on classification accuracy. In FIGURE 4a and d, speckle noise with a mean of 0 and variance of 0.09 was added to thermal image. This noisy image was classified by inception MV4, which erroneously identified it as cancerous with a probability of 99.5777%. This high misclassification in FIGURE 4a likelihood

highlights the vulnerability of the diagnostic model to noisy data, emphasising the importance of noise reduction techniques in ensuring reliable outputs.

After applying a denoising algorithm, FIGURE 4 b, c, e and f, represent the same noisy image. By removing the speckle noise, the diagnostic inception MV4 performance drastically improved. The denoised image was correctly classified, with an almost perfect confidence level of 99.99%. This remarkable improvement demonstrates the effectiveness of denoising filters in mitigating noise-related errors in medical imaging systems. The analysis underscores the critical role of pre-processing, such as noise reduction, in enhancing the reliability of AI-based diagnostic tools. Noise artefacts can significantly distort classification results, as evidenced by the nearly 100% cancer probability in the noisy image versus the accurate healthy classification after denoising. Such findings reinforce the necessity of robust pre-processing pipelines in sensitive applications like medical diagnostics.

3.5.2 Thermal image with Gaussian noise in different values

TABLE 9 and TABLE 10, evaluate the average detection accuracy of a model under Gaussian noise with a fixed variance of 0.02 and varying mean values (0.02 to 0.08), focusing on a misclassification scenario. The accuracy is measured over five detection attempts for each noise setting, and while the results show high performance, there is evidence of the noise affecting the model's accuracy, as seen in a slight drop to 99.99% during the fourth detection attempt when the mean is 0.05. This demonstrates that even small changes in Gaussian noise settings can significantly affect the deep learning model, leading to incorrect classifications and reducing its reliability in maintaining accurate detection under noisy conditions.

Table 9: Average detection accuracy of different values in variance of Gaussian noise

Setting (mean =0.05, variance is varied) (No Tipping Point)							
Accuracy in	0.02	0.03	0.04	0.05	0.06	0.07	0.08
1^{st} detection attempt	100	100	100	100	100	100	100
2^{nd} detection attempt	100	100	100	100	100	100	100
3^{rd} detection attempt	100	100	100	100	100	100	100
4^{th} detection attempt	100	100	100	99.99	100	100	100
5^{th} detection attempt	100	100	100	100	100	100	100
Average accuracy	100	100	100	99.998	100	100	100

Table 10: Comparison of parameters in Figure 6, such as Mean Noise Value, Standard Deviation of Noise and Noise Value in dB.

FIGURE 6	**Mean Noise Value**	**Standard Deviation of Noise**	**Noise Value in dB**
a	109.318	112.828	18.765
b	92.156	119.703	11.926
c	86.018	117.159	12.951

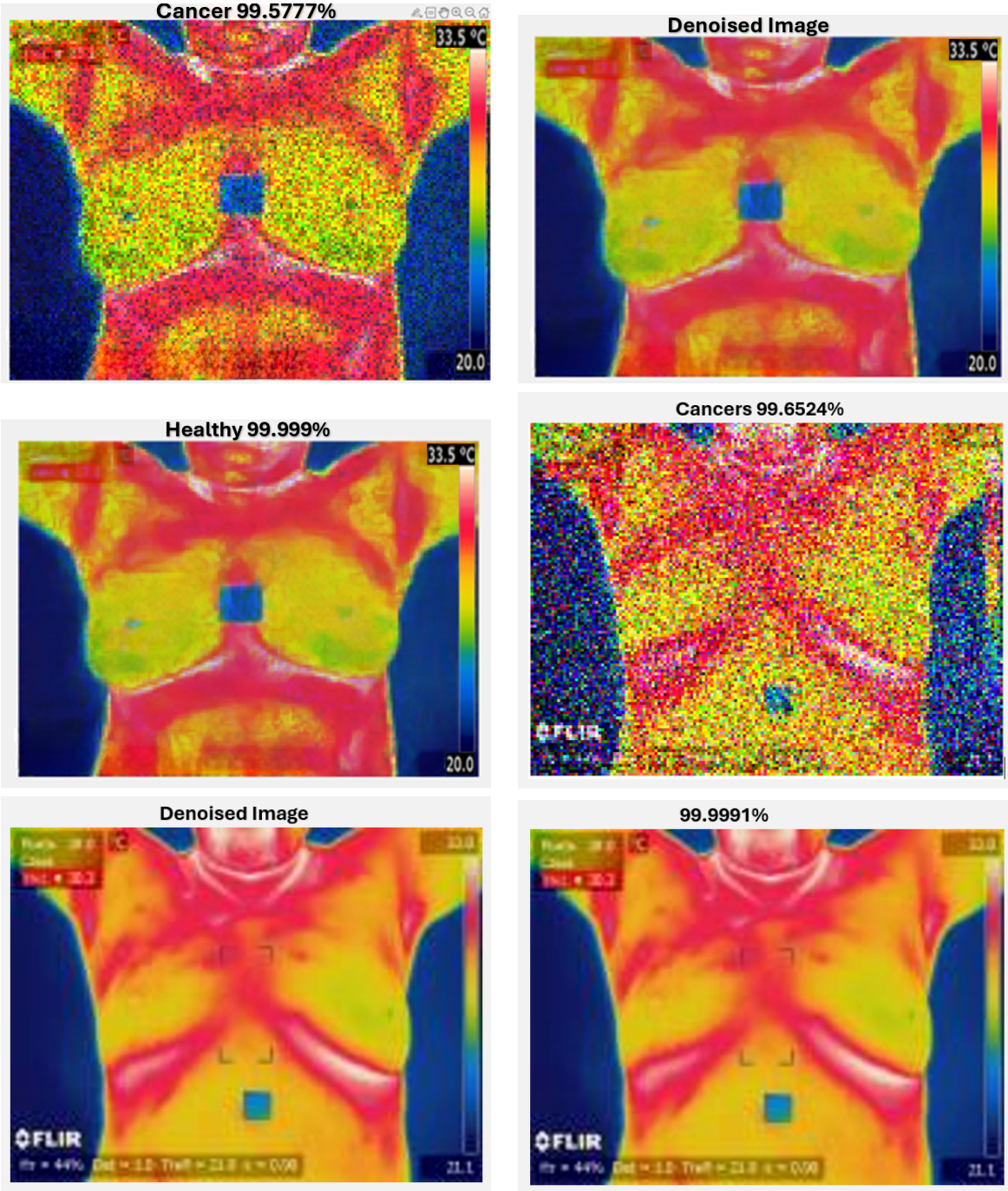

Figure 4: a) Healthy thermal image with speckle noise with a mean of 0 and variance of 0.09 was added, b) after applying a denoising algorithm, c) classified as healthy with 99.999 %. d) Cancer thermal image with speckle noise with a mean of 0 and variance of 0.09 was added, e) after applying a denoising algorithm, f) classified as Cancer with 99.9991%.

FIGURE 5 demonstrates the effects of Gaussian Noise on a thermal image and its subsequent denoising and classification using the Inception MV4 model. FIGURE 5a shows a thermal scan classified as "Cancer" with 100%. This classification error is due to the intentional addition of Gaussian Noise, which distorts the thermal data, making it unreliable for accurate diagnostic classification but FIGURE 5d classification was correct. The noise manifests as irregular multicoloured speckles that obscure the original thermal signature of the breast tissue. In FIGURE 5b and e, a denoising filter has been applied to reduce the Gaussian Noise. The results are evident in the smoother transition between colours and improved clarity of the thermal signature. Although the denoised image retains some residual artefacts from the noise, the filter substantially restores the visual integrity of the thermal data. FIGURE 5c represents the denoised thermal image, classified as "Healthy", with 100% accuracy by the Inception MV4 model. The thermal gradients are distinctly visible, with precise and continuous transitions in temperature. This demonstrates the model's ability to classify the thermal image accurately when free from noise distortion. The comparison highlights the robustness of the denoising process and its importance in mitigating misclassification caused by noisy data. It also underscores the significance of pre-processing in thermal imaging to ensure accurate AI-based diagnostic outcomes.

TABLE 10 provides a quantitative assessment of the noise levels present in the images, supporting the analysis with detailed metrics. For FIGURE 4a, the mean noise value is 109.318, with a standard deviation of 112.828 and a noise value of 18.765 dB, indicating the highest level of distortion among the images. For FIGURE 4b, the noise is reduced after applying the denoising filter, as evident from the mean noise value of 92.156, a standard deviation of 119.703, and a noise value of 11.926 dB. FIGURE 4c represents the clean thermal image with further reduced noise metrics: a mean noise value of 86.018, a standard deviation of 117.159, and a noise value of 12.951 dB. TABLE 10 demonstrates a clear relationship between noise characteristics and classification accuracy. As noise levels decrease through filtering, thermal patterns become more discernible, improving the model's diagnostic reliability.

3.5.3 Thermal image with salt & pepper noise with different values

TABLE 11 evaluates five individual accuracy measurements (accuracy 1 to accuracy 5) as well as an overall average accuracy. For Salt and pepper noise levels 0.1 and 0.2, all accuracy measures consistently report 100%, indicating that the detection system performs perfectly under low noise intensities. At 0.26 noise, a slight reduction is observed: accuracy 1 is 99.96%, accuracy 2 is 99.99%, accuracy three drops to 89.6%, accuracy 4 is 99.92%, and accuracy 5 is 99.93%, culminating in an average accuracy of 97.88%. As the noise increases to 0.27, the performance further decreases. The average accuracy drops to 94.876%, with individual accuracies between 85% and 99.8%. At a noise level of 0.28, the detection capabilities deteriorate further. The average accuracy declines to 71.06%, with a sharp drop in accuracy 1 to 62%, accuracy 3 to 99.9%, and accuracy 5 to 20%. Interestingly, performance improves slightly at 0.29, with the average accuracy increasing to 81.6%. Accuracy 1, 4, and 5 improve to 64%, 75%, and 81%. However, a significant degradation occurs at the highest noise level (0.3). The average accuracy plunges to 51.58%, with accuracy 1, 4, and 5 dropping to 18%, 52%, and 60%, respectively. This trend illustrates the system's sensitivity to high noise levels, emphasising its robust performance under low noise but substantial vulnerability as noise intensity increases.

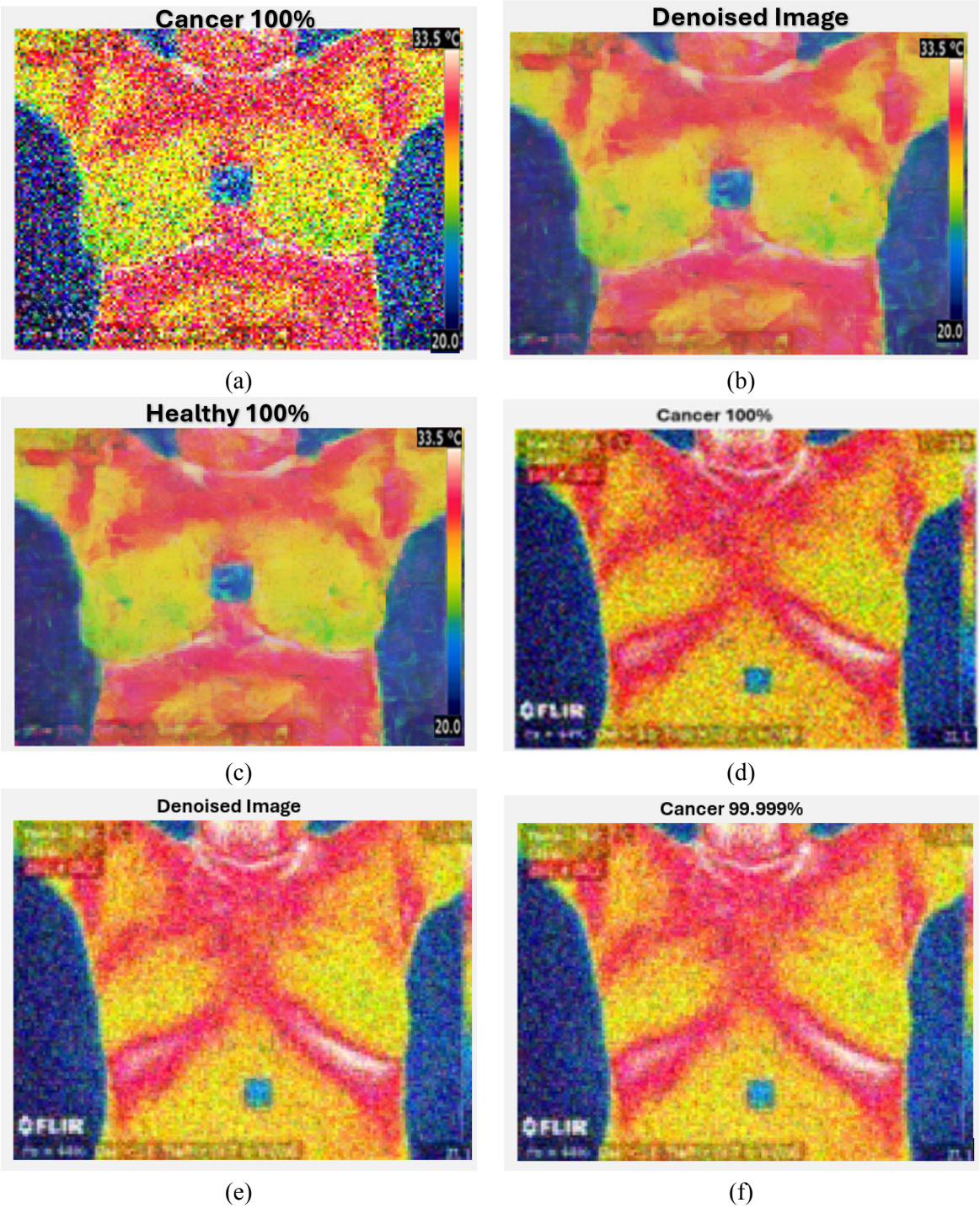

Figure 5: a) Healthy thermal image adding Gaussian Noise classified as Cancerous, b) after applying a denoising algorithm, c) classified as healthy with 100%. d) Cancer thermal image adding Gaussian Noise classified as Cancerous, e) after applying a denoising algorithm, f) classified as Cancer with 99.99%.

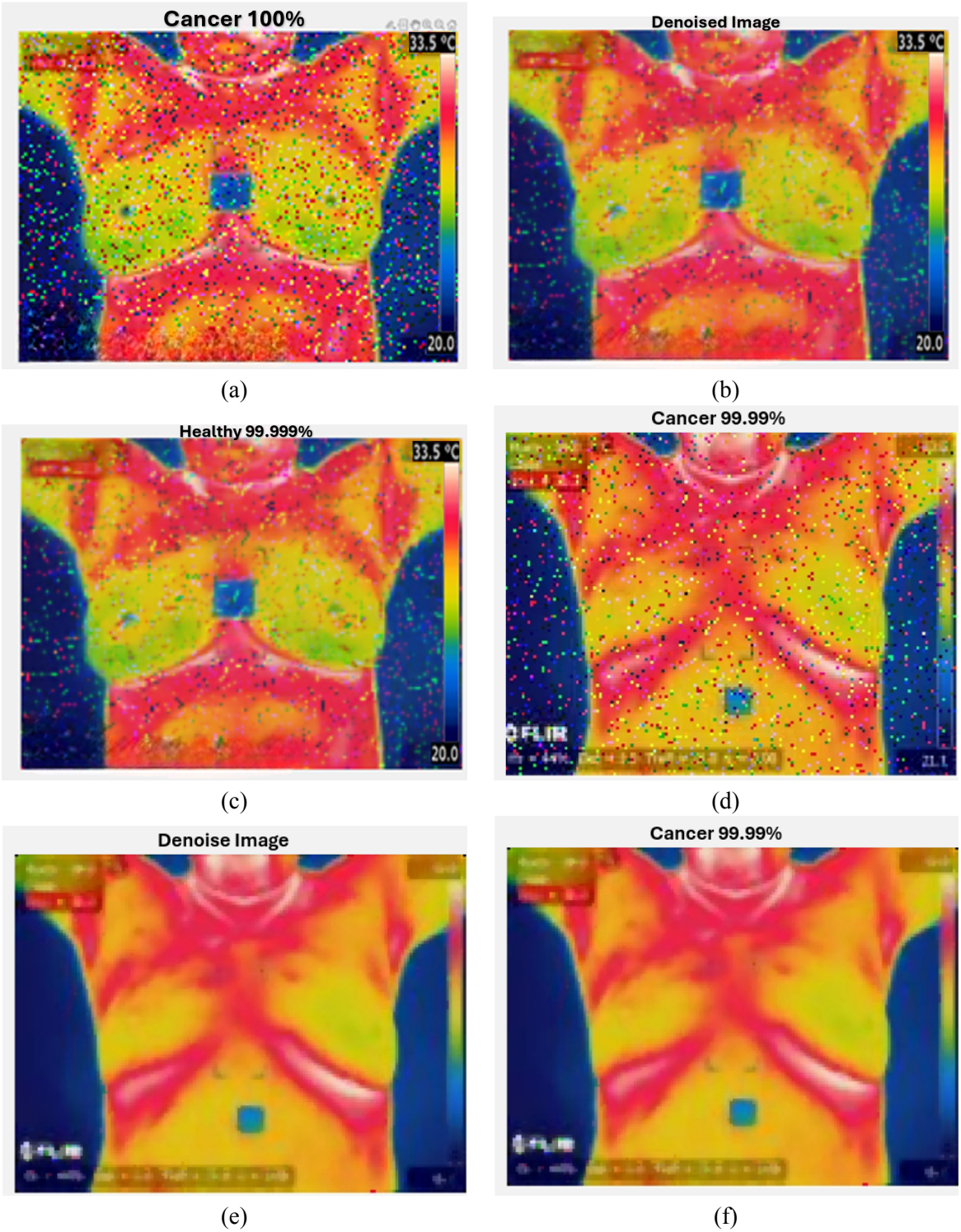

Figure 6: a) Healthy thermal images with adding salt & pepper noise classified as Cancer, b) Denoising thermal image, c) classified thermal images after removing the noise as Healthy 99.999%. d) Cancer thermal images with adding salt & pepper noise classified as Cancer, b) Denoising thermal image, c) classified thermal images after removing the noise as Cancer 99.99%.

Table 11: Average detection accuracy of different variance settings in Salt & Pepper noise (Detection Tipping Point @ variance = 0.3)

Setting (Mean=0, variance is varied)							
Accuracy in	0.1	0.2	0.26	0.27	0.28	0.29	0.3
1^{st} detection attempt	100	100	99.96	85	62	64	18
2^{nd} detection attempt	100	100	99.99	99.9	74	99	99.9
3^{rd} detection attempt	100	100	89.6	91.5	99.9	89	28
4^{th} detection attempt	100	100	99.92	98	99.4	75	52
5^{th} detection attempt	100	100	99.93	99.98	20	81	60
Average accuracy	100	100	97.88	94.876	71.06	81.6	51.58

FIGURE 6 illustrates the impact of adding salt and pepper noise to a thermal image of a breast and its classification results. The original thermal image has been manipulated by introducing salt-and-pepper noise with a mean of 0 and a variance of 0.09. FIGURE 6. a represents the noisy version of the healthy image, where the classification model predicts a cancer diagnosis with a level of 99.57%. This demonstrates the susceptibility of the classification model to noise, which significantly alters the outcome, but in FIGURE 6d, represents the noisy version of the cancer thermal image without any effect in classification. FIGURE 6b and e, and FIGURE 6e is a denoised version of the noisy image, where denoising techniques were applied to recover the original signal. However, residual noise or distortion may still affect the classification accuracy. FIGURE 6c and f, represents the thermal image state classification, where the original noise-free image is processed. As the variance increases, healthy images in the classification begin to flip being misclassified as cancer, with this shift occurring notably at a variance value of 0.3.

3.5.4 Thermal image with poisson noise with different values

TABLE 12 demonstrates the detection accuracy of inception MV4 when exposed to Poisson Noise at different levels, characterised by varying signal-to-noise ratio (SNR) values. Poisson noise, which often arises in low-light imaging or photon-limited applications, is applied to test the robustness of the detection algorithm. The table is structured to present accuracy values (percentage) across five test settings, along with the average detection accuracy for each noise level. For Poisson noise at 0.2 SNR, the system achieves 100% accuracy across all settings, indicating no performance degradation. At 0.3 SNR, a slight drop in performance is observed, with an average accuracy of 99.9%. For individual settings, the accuracy is consistently 99.9%, except for setting 3, where the SNR is recorded as 13.98 dB. At 0.4 SNR, the algorithm maintains a high performance of 100% accuracy across all settings, reflecting robustness at this noise level.

As the SNR decreases further, such as at 0.5 SNR, a minor drop in accuracy is observed in setting 3, where the SNR is noted as 10.46 dB, yet the overall average accuracy remains 99.9%. At 0.6 SNR, the system's performance in setting three drops slightly further to 6.02 dB, but the overall detection accuracy averages 99.9%. The degradation is more noticeable at 0.7 and 0.8 SNR, with setting 3 registering SNR values of 4.44 dB, 3.10 dB, and 1.94 dB, respectively. Despite this, the overall accuracy remains consistently high at 99.9%. The table underscores the system's high resilience to varying levels of Poisson noise. Although minor drops in SNR values occur in specific settings (e.g.,

Table 12: Average detection accuracy of different value Poisson noise (no tipping point is detected)

Poisson noise Setting for difference values of Variance (Mean = Variance)							
Accuracy in	0.2 (SNR = 13.98 dB)	0.3 (SNR = 10.46 dB)	0.4 (SNR = 7.96 dB)	0.5 (SNR = 6.02 dB)	0.6 (SNR = 4.44 dB)	0.7 (SNR = 3.10 dB)	0.8 (SNR = 1.94 dB)
1^{st} detection attempt	100	99.9	100	99.9	100	99.9	99.9
2^{nd} detection attempt	100	99.9	100	99.9	100	99.9	99.9
3^{rd} detection attempt	100	99.9	100	99.9	100	99.9	99.9
4^{th} detection attempt	100	99.9	100	99.9	100	99.9	99.9
5^{th} detection attempt	100	99.9	100	99.9	100	99.9	99.9
Average accuracy	100	99.9	100	99.9	100	99.9	99.9

setting 3), the average accuracy remains stable across all noise levels. This suggests the system is well-suited for environments where noise is prevalent, maintaining reliable detection performance even under challenging conditions. Finally, no tipping point is detected because the classification never reaches a state where thermal images are consistently misclassified.

FIGURE 7 thermal images illustrate a classification scenario where Poisson noise impacts the outcome of using thermal imaging. FIGURE 7a and d, Poisson noise is applied to thermal image with parameters of mean $\mu = 0$ and variance $\sigma 2 = 0.6$. This noise introduces artificial pixel intensity variations, mimicking measurement errors or environmental interference. After classification, the image, initially healthy, is misclassified as indicating cancer, with an accuracy of 99.962% as shown in FIGURE 7a, but in FIGURE 7d, is not effect. This result underscores the sensitivity of the classification system to noise, highlighting the potential for false positives when the image is corrupted. In FIGURE 7b and e, a denoising algorithm is applied to the same noisy image. This filter effectively reduces the noise, restoring the original intensity patterns and improving the image's quality. Following this processing, the denoised image undergoes reclassification. FIGURE 7c and f, DCNN inception MV4 accurately classifies. The comparison of the two outcomes, one before denoising and the other after demonstrates the critical role of pre-processing techniques in ensuring accurate diagnostics. As observed in the initial erroneous diagnosis, the classifier's reliability is compromised without noise reduction.

FIGURE 8 illustrates the detection performance for Poisson noise settings across different signal-to-noise ratio (SNR) levels. Higher SNR levels show better detection performance, with higher detection probabilities at lower false alarm probabilities.

The Inception v3, v4, and MV4 models achieve high accuracy in breast cancer detection using thermal images through architectural innovations such as multi-scale feature extraction via parallel convolutional filters (1x1, 3x3, 5x5) within inception modules, enabling the capture of both localized temperature variations and broader thermal patterns. Inception MV4's modifications—such as adding convolutional layers and increasing filter counts in "Inception B" modules—enhance sensitivity to subtle thermal anomalies linked to tumors. These models leverage preprocessing techniques (e.g., denoising filters, region-of-interest cropping) to mitigate noise artifacts like Gaussian or salt-and-pepper noise, ensuring robust input data. With Inception MV4 achieving 99.975% accuracy on clean datasets, their performance stems from balancing efficiency (e.g., reduced training time) and precision. Techniques like Grad-CAM (Gradient-weighted Class Activation Mapping) reveal

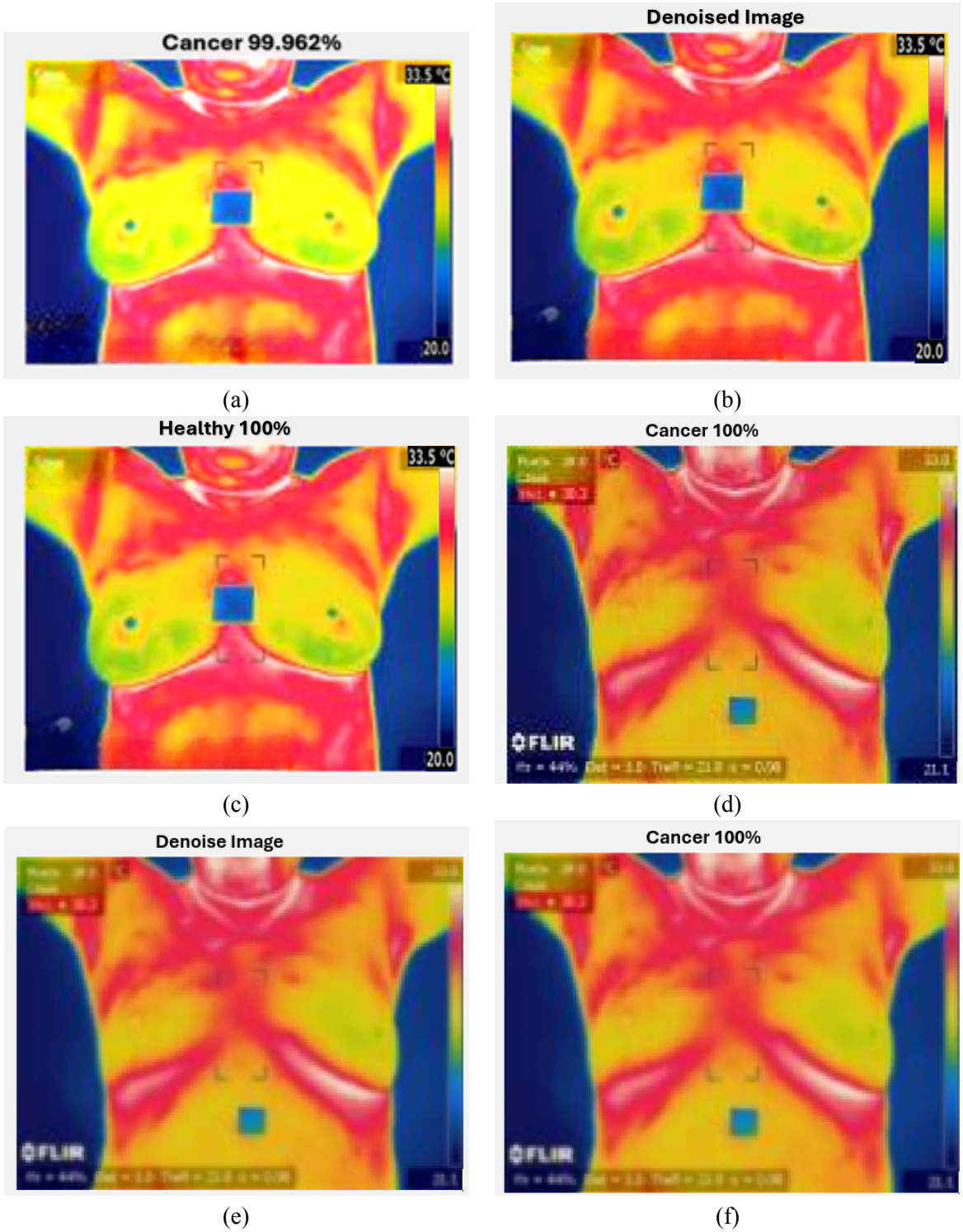

Figure 7: a) Healthy thermal image with adding Poisson noise classified as Cancer, b) Denoising thermal image, c) classified thermal images after removing the noise as Healthy 100%. d) Cancer thermal image with adding Poisson noise classified as Cancer, e) Denoising thermal image, f) classified thermal images after removing the noise as Cancer 100%.

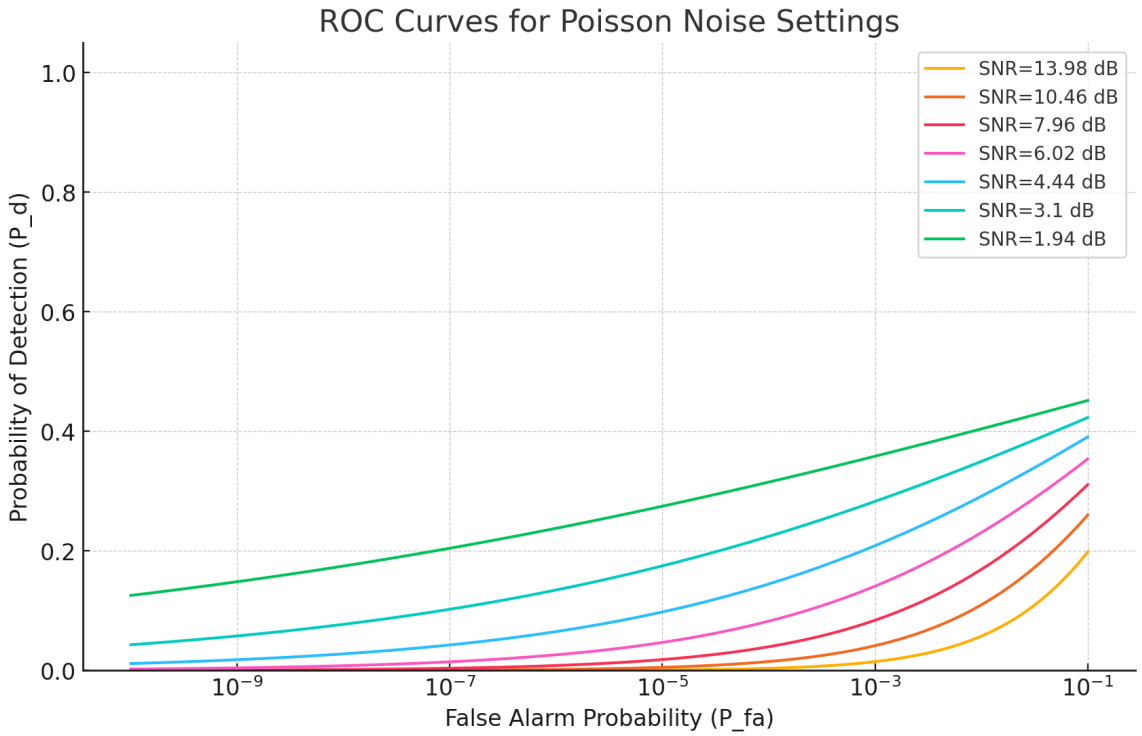

Figure 8: ROC Curves for Poisson Noise Settings at Various SNR Levels

that the models focus diagnostically relevant regions, such as asymmetric heat gradients or vascular patterns, by generating heatmaps that highlight areas like localized "hotspots." For instance, in noisy images misclassified as cancerous (e.g., FIGURE 4 in the study), Grad-CAM would show distorted thermal regions influencing errors, while denoising restores focus to normal patterns, aligning with corrected classifications. Feature visualization further validates that deeper layers prioritize textural and structural heterogeneity, such as edge detection via Sobel filters or metabolic heat variations. This interpretability underscores the models' reliance on clinically significant thermal signatures while exposing vulnerabilities to noise, as seen when salt-and-pepper noise (variance =0.3) mimics tumor-like artifacts, reducing accuracy to 51.58%. By integrating architectural robustness, noise mitigation, and explainability tools, these models advance reliable AI-driven diagnostics, fostering clinician trust and highlighting pathways for refining thermal imaging pipelines in early cancer detection.

TABLE 13, shows performance of Inception MV4 model on cancerous images under Poisson noise with SNR values ranging from 13.98 dB to 1.94 dB. At the highest SNR (13.98 dB), the model achieves perfect sensitivity and specificity (100%), reflecting ideal detection in low-noise conditions. As the SNR decreases, sensitivity gradually drops from 99.86% at 10.46 dB to 99.06% at 1.94 dB highlighting a slight impact of noise on detecting cancerous features. Despite this decline, specificity remains stable at 99.9%, and overall accuracy stays above 99.4% in all cases.

Table 13: Detection Accuracy under Poisson Noise at Various Signal-to-Noise Ratio (SNR) Levels on Cancerous Images

Poisson Noise Variance (SNR)	Sensitivity (Cancerous)	Specificity (Healthy)	Accuracy
0.2 (13.98 dB)	100%	100%	100%
0.3 (10.46 dB)	99.86%	99.9%	99.88%
0.4 (7.96 dB)	99.84%	100%	99.92%
0.5 (6.02 dB)	99.64%	99.9%	99.77%
0.6 (4.44 dB)	99.5%	100%	99.75%
0.7 (3.10 dB)	99.3%	99.9%	99.6%
0.8 (1.94 dB)	99.06%	99.9%	99.48%

4. CONCLUSIONS

The study was conducted on a dataset for classifying thermal images with added Gaussian Noise with different values to evaluate the influence of added noise on the performance of the newly introduced Deep CNN Inception MV4 in the classification of thermal images. The thermal image dataset must be free from any noise affecting classification accuracy. Salt noise had the most influence on the classification accuracy of the deep CNNMV4 without using filters during the pre-processing stage. However, adding noise filters are important in the pre-processing of thermal images before being fed into DCNN Inception MV4 as they mitigate the noise effect to a large degree. Moreover, the accuracy of classifying noise-free thermal images reached 99.974%, the highest accuracy compared to the rest of the noisy datasets. The model exhibited remarkable resilience to Gaussian noise, maintaining detection accuracy above 99.93% even with increasing noise levels. For speckle noise, detection accuracy remained at 100% for variances up to 0.04 but dropped significantly to 89.65% at 0.07 and 43.86% at 0.08, indicating sensitivity to higher noise intensities. Under salt-and-pepper noise, the model sustained perfect accuracy at lower noise levels (0.1 and 0.2) but experienced substantial performance degradation at 0.3, dropping accuracy to 51.58%. Poisson noise showed minimal impact, with an average detection accuracy of 99.9% across different signal-to-noise ratios (SNRs). Pre-processing techniques, such as denoising algorithms, were critical in mitigating noise-induced misclassifications. For example, noisy thermal images misclassified with up to 99.57% confidence were correctly identified after denoising, with classification accuracy improving to 99.999%. These results underscore the importance of effective noise reduction in ensuring reliable AI-based diagnostic outcomes. Although this study evaluates the impact of noise on thermal images of healthy individuals, it does not explore how noise affects the classification of thermal images with breast cancer. Future work will extend the analysis to include cancer-positive images to better understand how noise may alter or obscure malignant thermal patterns. Moreover, future work may explore the impact of incorporating regularization techniques, such as dropout and layer regularization, directly into the deep CNN architecture to enhance generalization performance and mitigate overfitting. Also, adding other types of noise to impact the image quality often experienced by stored and transferred thermal images to determine the best filters to process the noise before inserting the images into the DCNN. Moreover, proposed future research directions include developing representative datasets, integrating segmented images into the training process, and designing a lightweight (CNN) model to enhance CNN performance for application across various disease-related topics.

5. DECLARATIONS

5.1 Ethics approval and consent to participate

Not applicable

5.2 Consent for publication: Not applicable

5.3 Availability of data and materials

The datasets analysed during the current study are available in the DMR IR dataset repository, http://visual.ic.uff.br/dmi/

5.4 Competing interests

authors declare no competing interests

5.5 Funding

The authors are grateful to Arab Open University (AOU) for funding this research work.

5.6 Authors' contributions

MAS contributed to the paper by conducting the literature review, writing the manuscript, revising it, creating figures and tables, interpreting the results, and providing critical feedback and review.

MHH contributed to the paper by conducting the literature review, writing the manuscript, revising it, creating figures and tables, interpreting the results, and providing critical feedback and review.

MRI contributed to the paper by conducting the literature review, writing the manuscript, revising it, creating figures and tables, interpreting the results, and providing critical feedback and review.

YNS contributed to the paper by conducting the literature review, writing the manuscript, revising it, creating figures and tables, interpreting the results, and providing critical feedback and review.

References

[1] Hanf V, Kreienberg R. Corpus Uteri. 2020.

[2] Bini SA. Artificial Intelligence Machine Learning Deep Learning and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? J Arthroplasty.

2018;33:2358-2361.

[3] Yadav P, Jethani V. Breast Thermograms Analysisfor Cancer Detection Using Feature Extraction and Data Mining Technique. ACM Int Conf Proceeding Ser. 2016:1-5.

[4] Din NM, Dar RA, Rasool M, Assad A. Breast Cancer Detection Using Deep Learning: Datasets Methods and Challenges Ahead. Comput Biol Med. 2022;149:106073.

[5] Salvi S, Kadam A. Breast Cancer Detection Using Deep Learning and IoT Technologies. J Phys Conf Ser. 2021;1831:012030.

[6] Zhao P, Yoo I, Lancey R, Varghese E. Mobile Applications for Pain Management: An App Analysis for Clinical Usage. BMC Med Inform Decis Mak. 2019;19:1-10.

[7] Al Husaini MA, Habaebi MH, Hameed SA, Islam MR, Gunawan TS. A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. IEEE Access. 2020;8:208922-208937.

[8] Al Husaini MA, Habaebi MH, Islam MR, Gunawan TS. Self-Detection of Early Breast Cancer Application With Infrared Camera and Deep Learning. Electron. 2021;10:2538.

[9] Hiremath S, Karibasappa KG, Karibasappa K. Neural Network Based Noise Identification in Digital Images. ACEEE Int. J Netw Secur. 2011;02:28-31.

[10] Salami AM, Salih DM, Fadhil AF. Thermal Image Features and Noise Effects Analysis. In: Proceedings of the 7th international engineering conference research and innovation amid global pandemic. IEC Institute of Electrical and Electronics Engineers Inc. New York: IEEE. 2021:43-47.

[11] Liu Q, Liu Z, Yong S, Jia K, Razmjooy N. Computer-Aided Breast Cancer Diagnosis Based on Image Segmentation and Interval Analysis. Automatika. 2020;61:496-506.

[12] Priyadharsini MS. High Density Noise Filter Method for Denoising Mammogram Breast. Data cquisition Process. 2023;38.

[13] Sommer K, Plez B, Cohen-Tanugi J, Dagoret-Campagne S, Moniez M, et al. Stardice II: Calibration of an Uncooled Infrared Thermal Camera for Atmospheric Gray Extinction Characterization. Sensors. 2024;24:4498.

[14] Gade R, Moeslund TB. Thermal Cameras and Applications: A Survey. Mach Vis Appl. 2014;25:245-262.

[15] Wishart GC, Campisi M, Boswell M, Chapman D, Shackleton V, et al. The Accuracy of Digital Infrared Imaging for Breast Cancer Detection in Women Undergoing Breast Biopsy. Eur J Surg Oncol. 2010;36:535-540.

[16] Antony L, Arathy K, Sudarsan N, Muralidharan MN, Ansari S. Breast Tumor Parameter Estimation and Interactive 3D Thermal Tomography Using Discrete Thermal Sensor Data. Biomed Phys Eng Express. 2020;7:015013.

[17] Husaini MA, HABAEBI MH, HAMEED SA, ISLAM MR, GUNAWAN TS. A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. IEEE Access. 2020;8:208922-208937.

[18] Mulaveesala R, Dua G. Non-invasive and Non-ionizing Depth Resolved Infra-Red Imaging for Detection and Evaluation of Breast Cancer: A Numerical Study. Biomed Phys Eng Express. 2016;2:1-5.

[19] Yousefi B, Akbari H, Hershman M, Kawakita S, Fernandes HC, et al. SPAER: Sparse Deep Convolutional Autoencoder Model to Extract Low Dimensional Imaging Biomarkers for Early Detection of Breast Cancer Using Dynamic Thermography. Appl Sci. 2021;11:3248.

[20] Ekici S, Jawzal H. Breast Cancer Diagnosis Using Thermography and Convolutional Neural Networks. Med Hypotheses. 2020;137:109542.

[21] Kermani S, Samadzadehaghdam N, EtehadTavakol M. Automatic Color Segmentation of Breast Infrared Images Using a Gaussian Mixture Model. Optik. 2015;126:3288-3294.

[22] Dalmia A, Kakileti ST, Manjunath G. Exploring Deep Learning Networks for Tumour Segmentation in Infrared Images. 14th Quant InfraRed Thermogr Conf. 2018.

[23] Roslidar R, Syaryadhi M, Saddami K, Pradhan B, Arnia F, et al. Breacnet: A High-Accuracy Breast Thermogram Classifier Based on Mobile Convolutional Neural Network. Math Biosci Eng. 2022;19:1304-1331.

[24] Gomathi P, Muniraj C, Periasamy PS. Digital Infrared Thermal Imaging System Based Breast Cancer Diagnosis Using 4D U-Net Segmentation. Biomed Signal Process Control. 2023;85:104792.

[25] Goodman JW. Speckle Phenomena in Optics: Theory and Applications, 2nd ed. Bellingham, WA, USA: SPIE Press, 2020;PM312.

[26] Hou F, Zhang Y, Zhou Y, Zhang M, Lv B, et al. Review on Infrared Imaging Technology. Sustainability. 2022;14:11161.

[27] Robinson S. Editor. The Infrared & Electro-Optical Systems Handbook, Vol. 8: Emerging Systems and Technologies. Bellingham, WA, USA: SPIE Press. 1993.

[28] Stewart J. Human Medical Thermography. 2023.

[29] Raju PR, Hussain SA. A Computer-Aided Diagnosis Tool for Objective Assessment of Tumors Using Infrared Imaging. Int J Eng Res Comput Appl. 2014;3:10–16.

[30] Zmuidzinas J. Thermal Noise and Correlations in Photon Detection. Appl Opt. 2003;42:4989-5008.

[31] Besikci C. Nature Allows High Sensitivity Thermal Imaging With Type-I Quantum Wells Without Optical Couplers: A Grating-Free Quantum Well Infrared Photodetector With High Conversion Efficiency. IEEE J Quantum Electron. 2021;57:1-12.

[32] Al Husaini MA, Habaebi MH, Gunawan TS, Islam MR, Elsheikh EA, et al. Thermal-Based Early Breast Cancer Detection Using Inception V3 Inception V4 and Modified Inception MV4. Neural Comput Appl. 2022;34:333-348.

[33] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-V4 Inception-ResNet and the Impact of Residual Connections on Learning. 2016. ArXiv preprint:- https://arxiv.org/pdf/1602.07261v1.

[34] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-V4 Inception-ResNet and the Impact of Residual Connections on Learning (AAAI-17). AAAI. 2017;31:4278-4284.

[35] https://visual.ic.uff.br/en/proeng/thiagoelias/

[36] Goodman JW. Some Fundamental Properties of Speckle. J Opt Soc Am. 1976;66:1145.

[37] Hiremath PS, Akkasaligar PT, Badiger S, Gunarathne G. Speckle Noise Reduction in Medical Ultrasound Images. In: Gunarathne GP editor. Adv. break. ultrasound imaging. InTech. 2013;1:1-8.

[38] Sudha S, Suresh GR, Sukanesh R. Speckle Noise Reduction in Ultrasound Images by Wavelet Thresholding Based on Weighted Variance. Int J Comput Theor Eng. 2009;1:7-12/1793-8202.

[39] Buades A, Coll B, Morel JM. A Non-local Algorithm for Image Denoising. Proc IEEE Comput Soc Conf. Comput Vis Pattern Recognition. CVPR. 2005;2:60-65.

[40] Rohit V, Ali J. A Comparative Study of Various Types of Image Noise and Efficient Noise Removal Techniques. Int J Adv Res Comput Sci Softw Eng. 2013;3:2277–2128.

[41] Salmon J, Harmany Z, Deledalle CA, Willett R. Poisson Noise Reduction With Non-local PCA. J Math Imaging Vis. 2014;48:279-294.