Article

Evaluation of Integrated CNN, Transfer Learning, and BN with Thermography for Breast Cancer Detection

N. Aidossov [1], Vasilios Zarikas [2,3], Aigerim Mashekova [1], Yong Zhao [1], Eddie Yin Kwee Ng [4,*], Anna Midlenko [5] and Olzhas Mukhmetov [1]

[1] School of Engineering and Digital Sciences, Nazarbayev University, Astana 010000, Kazakhstan
[2] Department of Mathematics, University of Thessaly, GR-35100 Thessaly, Greece
[3] Mathematical Sciences Research Laboratory (MSRL), GR-35100 Lamia, Greece
[4] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore
[5] School of Medicine, Nazarbayev University, Astana 010000, Kazakhstan
[*] Correspondence: mykng@ntu.edu.sg

Abstract: Breast cancer comprises a serious public health concern. The three primary techniques for detecting breast cancer are ultrasound, mammography, and magnetic resonance imaging (MRI). However, the existing methods of diagnosis are not practical for regular mass screening at short time intervals. Thermography could be a solution to this issue because it is a non-invasive and low-cost method that can be used routinely as a self-screening method. The research significance of this work lies in the implementation and integration of multiple different AI techniques for achieving diagnosis based on breast thermograms from several data sources. The data sources contain 306 images. The concept of transfer learning with several pre-trained models is implemented. Bayesian Networks (BNs) are also used to have interpretability of the diagnosis. A novel feature extraction from images (related to temperature) has been implemented and feeds the BNs. Finally, all methods and the classification results of pre-trained models are compared. It is found that the best result amongst the transfer learning concept is achieved with MobileNet, which delivered 93.8% accuracy. Furthermore, the BN achieves an accuracy of 90.20%, and finally, the expert model that combines CNNs and BNs gives an accuracy of 90.85%, even with a limited amount of data available. The integration of CNN and BN aims to overcome the hardship of interpretability. These approaches demonstrate high performance with added interpretability compared to previous works. In conclusion, the deep neural network provides promising results in breast cancer detection. It could be an ideal candidate for Breast Self-Exam (BSE), the goal recommended by WHO for mass screening.

Keywords: breast cancer; thermography; Bayesian networks; convolutional neural network; transfer learning; expert model

1. Introduction

One of the most serious health issues that might have devastating effects for women nowadays is breast cancer. Changes in the cell genome, hormonal malfunction, family history, hormone therapy, lifestyle characteristics, and unfavorable life practices are potential risk factors and causes [1]. Breast cancer is a complex illness, the progression of which is linked to alterations in a cell's DNA brought on by environmental factors and hormones [2]. It is regarded as one of the illnesses that kills more women than any other. The rate of recovery varies depending on the stage of the disease when it is diagnosed. Therefore, early diagnosis is vital, such that the tumor is likely to be treated at earlier stages.

There are many cancer diagnosis techniques, and mammography is the gold-standard. Breast tissue conditions including density, past surgery, lactation, breast implantation, and hormones, which should be in a normal state during diagnosis, limit the accuracy of mammography, which is an X-ray image of the breast. One adverse effect of this technique

is that there could be harmful radiation, which is why mammography isn't recommended more than twice a year [3–8].

Another method for cancer diagnosis and breast screening is ultrasound. To determine if a lump is a solid tumor or a cyst filled with fluids, an ultrasound is advised as a preliminary test. Younger people can benefit from it because their mammary glands have a denser structure. However, the expertise of the medical professional doing the test determines whether ultrasonography diagnosis will be successful (note: its accuracy is a function of volume to mass ratio) [3–8].

Furthermore, a widely used method of breast cancer diagnosis and screening is MRI (magnetic resonance imaging). The most reliable and accurate way to diagnose a tumor or breast cancer is via MRI scans [3–8]. However, currently, this procedure is the most expensive and is only available in major, well-equipped hospitals. F-FDG PET/CT is an emerging detection technique that combines PET and CT, which has synergistic advantages over PET or CT alone and minimizes their individual limitations. However, due to excessive cost and limited availability, F-FDG PET/CT application is severely constrained. PET itself is also a rather expensive technology, and neither PET scanners nor the cyclotrons required to create isotopes for PETs are generally accessible. For the overall evaluation of breast lesions, SPECT imaging has a higher diagnostic value than mammography, has been extensively validated, and has a high sensitivity. However, SPECT is much more expensive and less widely available than other techniques [3–8].

Finally, there is a method called contrast-enhanced spectral mammography (CESM), which can be used to detect breast cancer and provides low-energy mammographic images comparable to standard digital mammography, as well as a post-contrast recombined image to evaluate tumor neovascularity, like magnetic resonance imaging (MRI). This technique, however, has limitations, such as the usage of an iodinated contrast agent, and increased radiation exposure [7].

Thus, every technique has its own advantages and disadvantages. The main disadvantages of the considered methods are that they are not practical for regular mass screening frequently, and some of them have limited access for people who live in remote areas of a country.

Thermography is one of the non-invasive and affordable techniques for routine and bulk screening [5,6]. It is common knowledge that many health problems can be accurately detected by a person's body temperature. Blood perfusion, metabolic rate, and ambient temperature are just a few examples of the variables that affect how hot or cold the body is. Thermography could detect any temperature abnormality in the body, such as a tumor, as most tumors cause temperature changes in the surrounding tissue [5–8].

Thermography is a type of imaging that uses infrared (IR) light to create colored images of temperature distributions. According to the studies [8] conducted in this field, the surface of a breast with cancerous tissues has a higher temperature profile than the surrounding region, and abnormalities can be discovered through thermography. The diagnosis of thermography is reliant on qualitative principles and human judgment, such as the asymmetry of two breasts, hyper-hermetic patterns, and atypical vascular patterns, even though thermography has straightforward functioning principles [9]. By examining the temperature distribution and randomly matching the temperature profiles at several locations, quantitative information is often manually derived. With the quick advancement of computer technology, computer-aided tools can be used to help with the diagnosis by supporting the interpretation of thermal images, creating better breast models, and automatically identifying the locations and sizes of tumors, as well as other unique characteristics of breast tissues [10].

The convolutional neural network (CNN), a deep-learning neural network of neurons with learnable weights and biases, is one newly discovered technique for recognizing images. The practice of identifying images based on their visual content is known as image classification. Recognizing breast thermograms with a predefined label is a crucial step in the learning process for neural networks. This is referred to as supervised learning.

Without any human expert intervention, a diagnostic tool can be built using CNN and CNN-based models to classify "healthy" and "sick" types of thermograms as diagnosed by doctors. This study examines various CNN-based models with varying parameters and develops an efficient CNN model for binary classification using breast thermograms.

In the recent developments in the usage of state-of-the-art and transfer learning for medical images, several studies have shown results in breast cancer thermography. For instance, in the study by Zuluaga-Gomez [11] they achieved an accuracy of 92% by using CNN with fifty-seven patients available data. Additionally, research by Torres-Galvan [12] demonstrated an accuracy of 91.2% by exploiting the VGG-16 model, with 173 patients in total. In the current study, several techniques that are not used in those papers are applied and try to outperform the results of the previous studies. Another advancement in the field is the development of Bayesian Networks. Bayesian Networks (BN) or influence diagrams are knowledge representation schemes capable of expressing every type of knowledge: discrete or continuous, certain or uncertain. Technically speaking, BNs are probabilistic networks between statistical factors with additional information about their causal/influence interconnections. BNs are an artificial intelligence method with interpretability. The latter means that the diagnosis is not coming from a black box. The physician will be able to understand, which is the crucial factor that results in the specific decision.

BNs are applied successfully to a large spectrum of different domains [13–18]. Perhaps, the most outstanding performance and meaningful use of BNs is for medical diagnosis. Almost perfect diagnostics were demonstrated many years ago [19]. The advantage of BNs lies in the fact that since they are probabilistic networks that connect statistical factors (random variables), they are interpretable. BNs are a knowledge representation scheme that encapsulates certain and uncertain knowledge. Therefore, when the network is trained by data (supervised learning), the result is not the estimation of weights on various layers but conditional probabilities among meaningful statistical factors. Thus, there is transparency and interpretability for any kind of diagnosis or probabilistic inference with BNs.

The aim of this paper is to assess several state-of-the-art neural network techniques for diagnosis from thermal images and compare their performances with those of the BNs. In addition, this study aims to demonstrate that by integrating BNs and CNN (or similar neural network algorithms), an expert system with remarkably high accuracy and, at the same time, interpretability with limited datasets can be generated. Furthermore, an automatic crop of the thermal images and automatic extraction of temperature-related features are implemented in the study. The latter is particularly important for BNs to achieve accuracy like CNNs. If the feature extraction is not appropriate, then BNs are not able to compete with CNNs. Therefore, this paper gives an answer to the obvious question, "when is a feature extraction scheme from images acceptable, and when can it be used for an interpretable knowledge representation BN structure"?

The current paper focuses on developing CNNs based on a multi-source database without preprocessing for binary classification. Furthermore, the paper compares transfer learning methods with the baseline CNN model to develop an intelligent tool for breast cancer detection. The paper further discusses the concepts of transfer learning and studies different models with this concept. Then it introduces the idea of Bayesian Networks and the data used, its implementation details are described, and an integrated BN + CNN model is presented. Finally, the results of BN and the integrated BN + CNN models are presented and discussed. In the end, a conclusion and discussion are given.
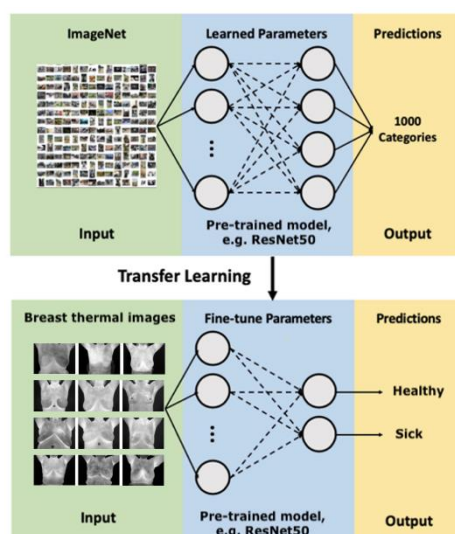
2. Materials and Methods

2.1. Classification of Images Using Transfer Learning

Transfer learning is a machine learning method where a model that has already been trained is used as the basis for a new model and task. In addition to requiring less effort to obtain training data, it can also accelerate the training procedure.

The current work has employed four pre-trained models for transfer learning. These pre-trained models were trained on the ImageNet dataset, which contains regular images of everyday objects. With the transfer learning technique, CNN models can be modified to recognize images in different formats.

ImageNet [20], a massive dataset with 1.2 million high-resolution pictures organized into 1000 different groups, was explored to pre-train models (e.g., fish, bird, tree, flowers, sport, room, etc.). Additionally, pre-trained model designs were altered: the last fully connected layer's output count was decreased from 1000 to the two classes required for each classification task. The refined model can classify input images into the class with the most significant score by producing scores for each class. A schematic representation of transfer learning with a set of models is presented in Figure 1.

Figure 1. Schema of transfer learning technique with pre-trained models for binary classification.

2.1.1. Very Deep Convolutional Networks (VGG16)

The VGG16 [21] is a well-known pre-trained model for image classification. The VGG16 model comprises convolutional, max-pooling, and fully connected layers. The total is 16 layers with five blocks, each with a max pooling layer of 16. With deep layers, VGG16 achieves excellent performance in the image competition.

2.1.2. Xception

F. Chollet [22] created a CNN neural network called Xception that was solely dedicated to depth-wise separable convolution layers. The network's foundation in this design for feature extraction is made up of 36 convolutional layers. The Keras Applications module was backed by an open-source Xception implementation utilizing Keras and TensorFlow under the MIT license [23]. On the ImageNet dataset, this architecture performs better than Inception V3, ResNet-50, ResNet-101, ResNet-152, and VGG. This model was among the primary choices because it was well-developed and produced outstanding results in various classification cases for small-sized datasets [22].

2.1.3. ResNet50

Deep convolutional networks known as residual networks (ResNets) use shortcut connections to create residual blocks to skip blocks of convolutional layers. Convolutional layers with a stride of two directly do the down sampling, and batch normalization is carried out immediately following each convolution and prior to ReLU activation. The identity shortcut is used when the input and output have identical dimensions. The projection shortcut matches dimensions through 11 convolutions as the dimensions rise. Both times, the shortcuts are executed with a stride of two when traversing feature maps of different

sizes. A 1000 completely connected layer with SoftMax activation makes up the network's last layer. There are 50 weighted layers, and there are 23,534,592 trainable parameters [24].

2.1.4. MobileNet

MobileNet uses depth-wise separable convolutions. From two operations, a depth-wise separable convolution is created: convolution in depth, and convolution in points. This provides a great starting point for training our classifiers, which are extremely small and super quick. Google open-sourced MobileNet, a type of CNN [25].

2.2. Statistical Methods and Construction of BN + CNN Expert Model from Data

The proposed methodology consists of the following steps:

1. Feature extraction from the thermal images. This step is described in more detail in another subsection. The extracted factors from the image are:
2. Maximum temperature;
3. Minimum temperature;
4. Max temperature minus min temperature;
5. Mean temperature;
6. Median temperature;
7. Standard deviation;
8. Variance;
9. Max temperature minus mean temperature;
10. Max temperature minus max temperature of the healthy breast;
11. Max temperature minus min temperature of the healthy breast;
12. Max temperature minus mean of the healthy breast;
13. Mean temperature minus mean temperature of the healthy breast;
14. Distance (max to min in pixels);
15. A = number of all pixels around the point of the maximum temperature that have temperature > [mean + 0.5 (maximum-mean)];
16. B = number of pixels/cells of the temperature matrix of the image of the breast with tumor;
17. C = number of all pixels that have temperature > [mean + 0.5 (maximum-mean)];
18. A/B;
19. C/B.
20. First, a descriptive analysis must be performed to check the distributions and the probability density functions. Then, appropriate algorithms will be used for the imputation of missing values. In our case, the structural expectation maximization algorithm as it is implemented in BayesiaLab 10.2 software was used;
21. BNs require the discretization of scale variables. A supervised multivariate discretization method optimized for the target variable tumor (positive = 1 or negative = 0 states) has been used;
22. Unsupervised learning has been performed to discover associations and possible causal structures. This step is independent of the supervised learning that finally will be used to build the final expert model. However, it is a consistency check and can also provide the influential nodes for the target variable, which is the tumor. Arrow-type connections and strengths of informational nodes are evaluated with the information-theoretic mathematical quantities of Mutual Information, Minimum Description Length (MDL) score, and the Kullback-Leibler Divergence or KL Divergence;
23. Supervised learning was performed using the Augmented Naïve Bayes learning scheme. Next, validation of the results was checked using K-fold analysis;
24. Run the supervised learning for structuring a BN using the extracted features plus the following historical medical data for each patient: marital status, race, age, first menstruation, last menstrual period, eating fat, mammography, radiotherapy, plastic surgery, prosthesis, biopsy, hormone replacement, wart on breast and temperature;

25. Run the same supervised learning for creating BN with the same data plus CNN prediction factor;
26. Compare. If the accuracy is similar, then accept the integrated CNN + BN expert model that achieves both explainability/interpretability and high accuracy. It is important for BN to have similar performance with the CNN methods because this means that the influential factors that can be discovered only with the BN method (and are significant for a physician) are enough for a good diagnosis.
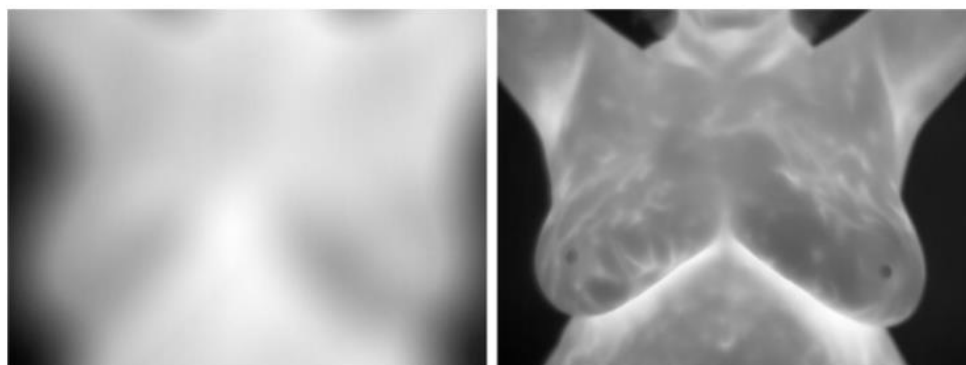
3. Experimental Setup and Results

3.1. Dataset Description

The dataset for this study consisted of 306 thermal images. The presented research was certified by the Institutional Research Ethics Committee (IREC) of Nazarbayev University (identification number: 294/17062020). To create a consolidated dataset, images were collected from two different datasets. The first source of data was the Database for Mastology Research [26,27], which is publicly available and managed by overseas researchers. The second was obtained locally by the researchers on this ongoing project in the Multifunctional Hospital of Astana.

3.1.1. The Database for Mastology Research

The Database for Mastology Research contained 266 thermal images, which were extracted as input for our diagnosis tool [26,27]. During the data collection stage, several images were excluded for usage in this research for the following reasons:

- Images are fuzzy, with the contour of the breasts and even the entire body barely apparent. As can be seen in Figure 2, the distinction is quite noticeable.
- Images that did not adhere to the approved protocol, for instance, nineteen patients had their arms down or were photographed in an odd position.

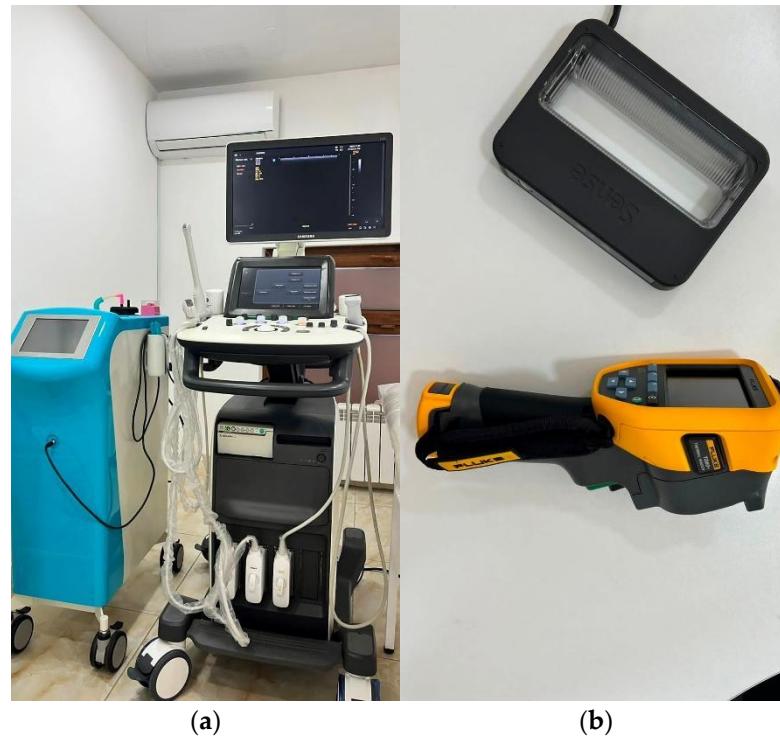**Figure 2.** Depiction of fuzzy (**left**) and accepted (**right**) thermal images of the database.

3.1.2. The Multifunctional Medical Center of Astana

The Multifunctional Medical Center of Astana gathered thermograms of patients who consented to collect thermal breast images. For this project, forty thermal images with human features were selected. Breast thermograms for women aged 18 to 80 are currently available in the database. Figure 3 shows the clinical office and IR camera FLUKE TiS60+ used for collecting thermograms.

A summary of the consolidated dataset is given in Table 1, and the sample grayscale thermal images are given in Figure 4.

The combined dataset consisted of a total of 306 images; 123 thermal images of patients classified as "sick" (as indicated in the doctor's diagnosis report), and 183 thermal images classified as "healthy". Patients' thermographic images classified as "healthy" represented that the patient was not diagnosed with having tumors or any cancerous cells in the breast area. If the thermograms were labeled as "sick," the patient had a tumor. It may have been either malignant or benign. The dataset was be divided into three sets dedicated to training,
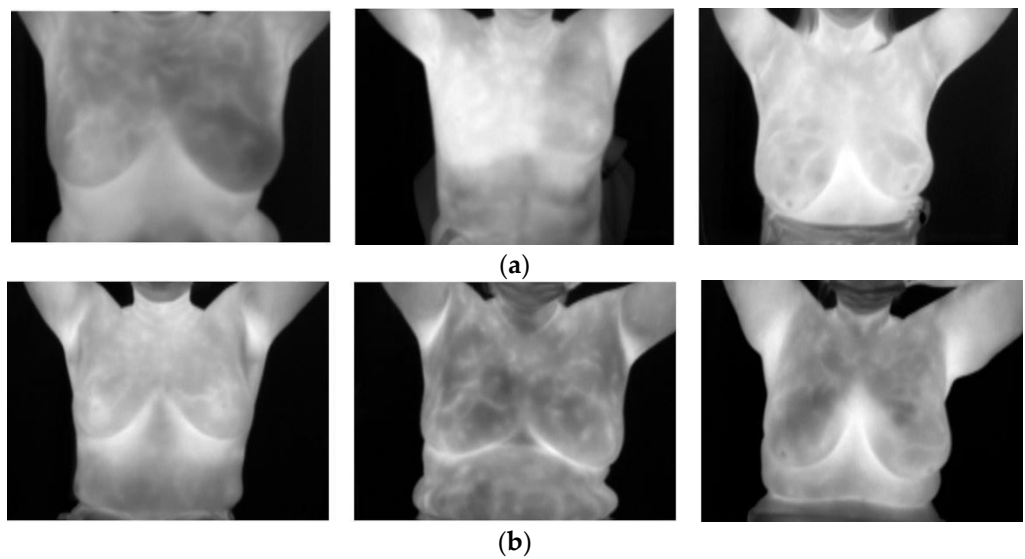
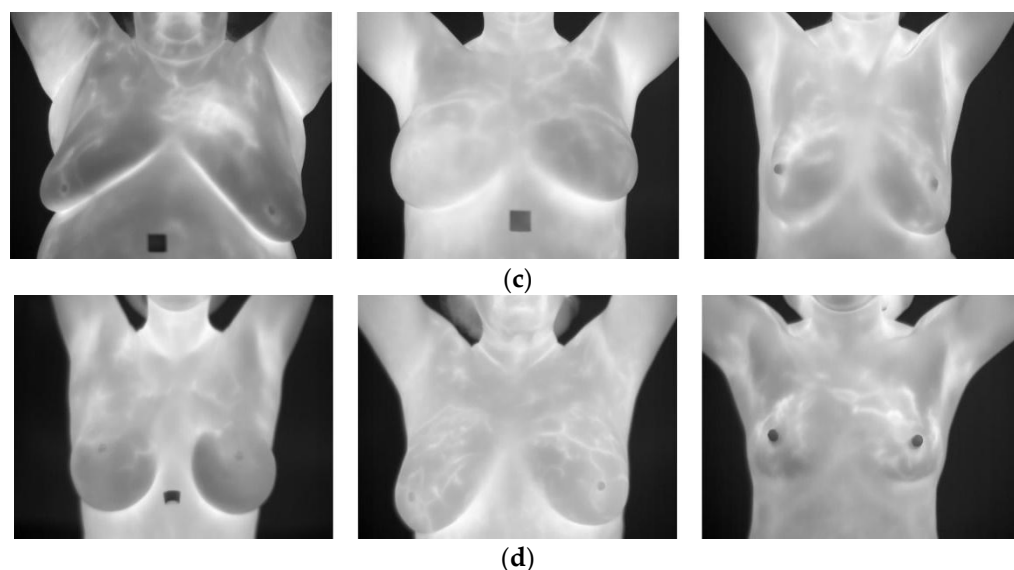cross-validation, and testing purposes. Table 2 reveals the details of our final training and testing sets.

(**a**)   (**b**)

Figure 3. Equipment and space where thermograms from patients were retrieved (**a**) Clinical office for collecting patient data; (**b**) 3D Sense Scanner (on top) and IR camera FLUKE TiS60+ (bottom).

Table 1. Details of the dataset.

	Healthy	**Sick**	**Total**
Database for Mastology Research (DMR)	166	100	266
Multifunctional Medical Center Astana	12	28	40
Total	178	128	306

(**a**)

(**b**)

Figure 4. *Cont.*

Figure 4. Sample thermal images of breasts from the dataset: (**a**) sick (Hospital of Nur-Sultan), (**b**) healthy (Hospital of Nur-Sultan), (**c**) sick (DMR), and (**d**) healthy (DMR).

Table 2. Details of training, cross-validation, and testing dataset for the classification model.

Label	Training	Cross-Validation	Testing	Total
Healthy	122	18	38	178
Sick	90	12	26	128
Total	212	30	64	306

3.2. Data Augmentation and Implementation Details

The generalization issue with deep learning methods was addressed by data augmentation, which enables the network to learn variations and generalize its findings to previously unobserved images while preventing overfitting on small datasets. This method has already been shown to be effective for classifying medical images [28]. So, the following operations served as the direction for each image:

- Rotation: rotating the image with an angle between zero and ten in the clockwise or counterclockwise direction;
- Scaling: sampling the scale of the frame size of the image randomly between 80% and 110%;
- Translation: translating the image horizontally and vertically between −10% and 10%;
- Horizontal flip: horizontally flipping the image with a probability of 0.5.

For this study, all thermal pictures were normalized and resized. In the testing stage, the model and tune of the hyperparameters were evaluated, and generalized results were obtained. For training, the following hyperparameters are used: Batch size = 32; Optimizer = "adam"; epochs = 25. Batch size is the number of images that are processed in one iteration, "adam" is the type of optimizer which is an extension to the Stochastic Gradient Descent, and epoch is passing the data forward and backward one time in a neural network. The first callback list was also defined as follows. The Early Stopping function had to be determined for efficient learning. To avoid overfitting, this function was used. The "loss" value is tracked in this CNN model, and the "patience" of the three epochs is found. It means that once the loss value reaches a minimum, and if the loss value increases in the next three epochs, training will be terminated at that epoch. Another adherence strategy is to slow down the learning rate. As a result, once the metric reaches a plateau, the learning rate slows. For this callback, patience is set to two epochs. If no improvement is detected, the learning rate is reduced by a factor of 0.3 because the loss value will gradually decrease until it reaches its lowest value.

Class weight was another critical parameter to define. Because the dataset primarily consisted of patients who did not have breast cancer, it was necessary to give minority classes a higher-class weight so that they could learn from all classes in a balanced manner [29].

The study was performed on a machine with the characteristics given in Table 3.

Table 3. Characteristics of the machine where computation was conducted.

Component	Value
CPU	Intel ® Xeon ® Silver 4210 Processor
Total Cores	10
Total Threads	20
Max Turbo Frequency3	3.20 GHz
Processor Base Frequency2	2.20 GHz
Cache	13.75 MB
RAM size	64 GB
Maximum Memory Speed	2400 MHz

3.3. Evaluation Metrics

Eight evaluation metrics are used to evaluate the models because of data imbalance problems, namely: *accuracy*, *precision*, *recall* (*sensitivity*), *specificity* (or *selectivity*), *F1-score*, confusion matrix, ROC (receiver operating characteristic) curve, and the area under the curve (AUC).

On a classification task, a confusion matrix provides a summary of the predicted and true labels.

Accuracy, *precision*, *recall*, *specificity*, and *F1-score* can be calculated by the following formulas:

$$accuracy = \frac{TP + FN}{TP + FP + TN + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall \; (or \; sensitivity) = \frac{TP}{TP + FN} \tag{3}$$

$$specificity \; (or \; selectivity) = \frac{TN}{TN + FP} \tag{4}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{5}$$

ROC curve is a graphical representation of a binary classification model's performance. It is determined by comparing the true positive rate (*TPR*) to the false positive rate (*FPR*) at various discrimination thresholds, where *TPR* is also known as sensitivity or recall and *FPR* is known as false positive rate (1-*specificity*). By adjusting the model's discrimination threshold, the confusion matrix can be changed, and a point on the ROC curve can be plotted.
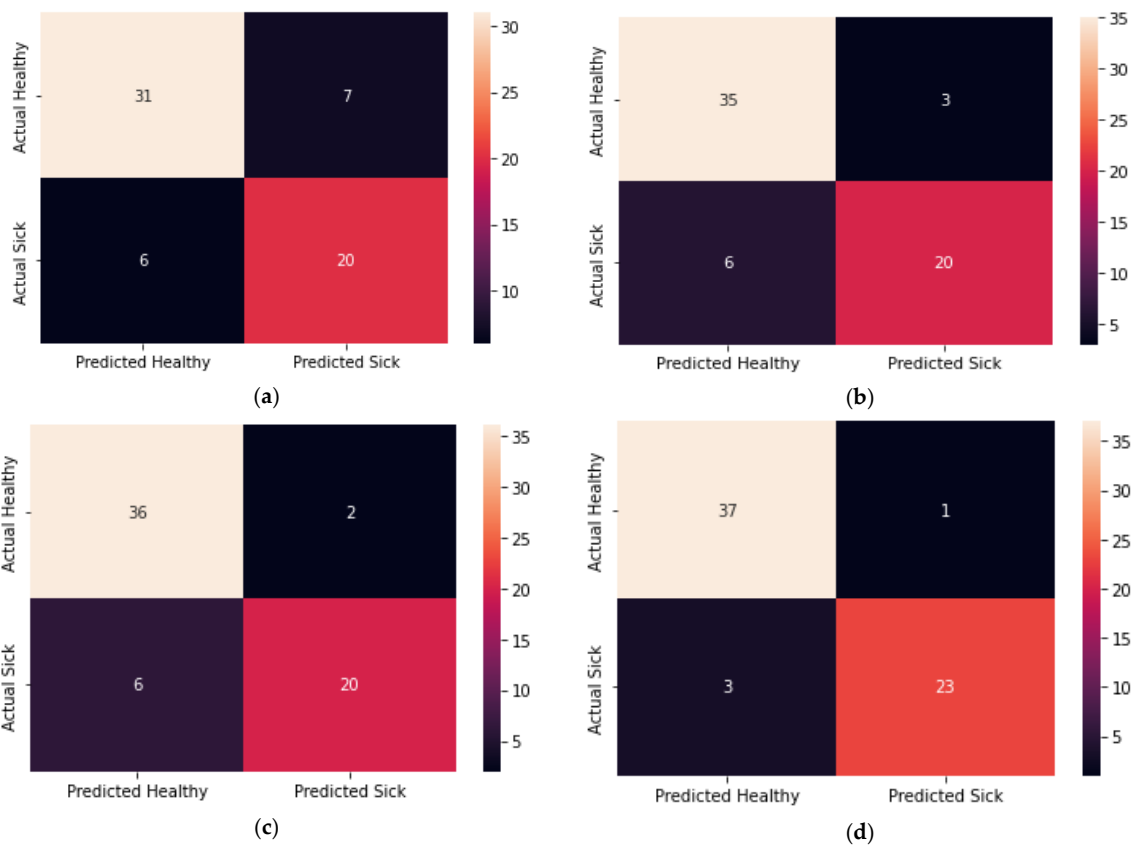
3.4. Results for Image Classification

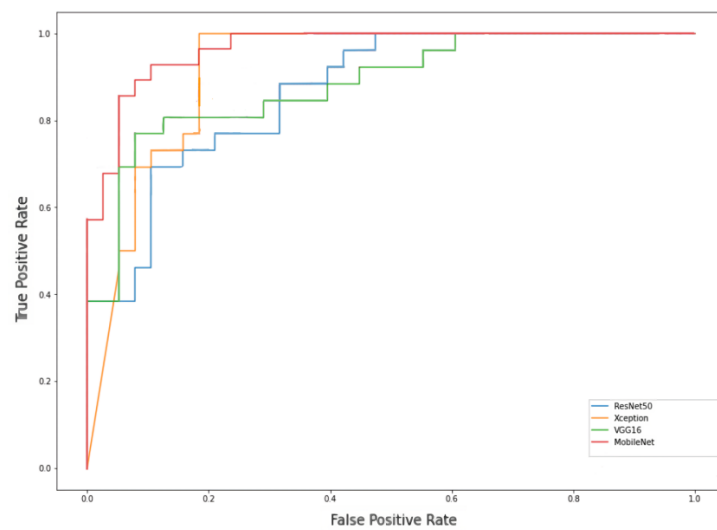3.4.1. Transfer Learning Models' Results

Four different models for this study generated comparatively high results with the transfer learning approach. Resnet50, VGG16, Xception, and MobileNet achieved an accuracy of 90.6%, 92.2%, 85.6%, and 93.8%, respectively. Table 4 summarizes the details of the results of each of the pre-trained models and baseline CNN model from the previous subsection in terms of *accuracy*, *precision*, *recall*, *f1-score*, *selectivity*, and AUC score. Figure 5 visualizes the confusion matrix for each model in binary classification. Finally, Figure 6 depicts the ROC curve for pre-trained models.

Table 4. Performance metrics juxtaposition.

CNN Model	Accuracy	Precision	Sensitivity/Recall	Specificity	*F*1-Score	AUC Score
Xception	85.9	85.4	92.1	76.9	88.6	0.92
MobileNet	93.8	92.5	97.4	88.5	94.9	0.974
ResNet50	79.7	83.8	81.6	76.9	82.7	0.967
VGG16	87.5	85.7	94.7	76.9	90.0	0.969

Figure 5. Confusion matrix of transfer learning approach models (**a**) ResNet50 (**b**) Xception (**c**) VGG16 (**d**) MobileNet.

Figure 6. ROC curve for all models.

3.4.2. Bayesian Network Results

In this work, the results from the constriction of two Bayesian base expert models are reported, one that uses information from the images plus medical record data, and a second one that also includes diagnostic information from a CNN model.

Supervised learning performed with random sampling was used for training 30% of the data for testing, and 70% was used as a learning set. BayesiaLab 10.2 software was used for the calculations [14].

Figures 7–13 show that the BN + CNN expert model is consistent. The most crucial factors for the diagnosis decision are min temperature, max temperature, abs (max-max healthy), biopsy, race, A/B, and CNN prediction. Tables 5 and 6 demonstrate particularly superior performance, similar to CNN methods. The accuracy of the BN + CNN expert model was 90.85%.
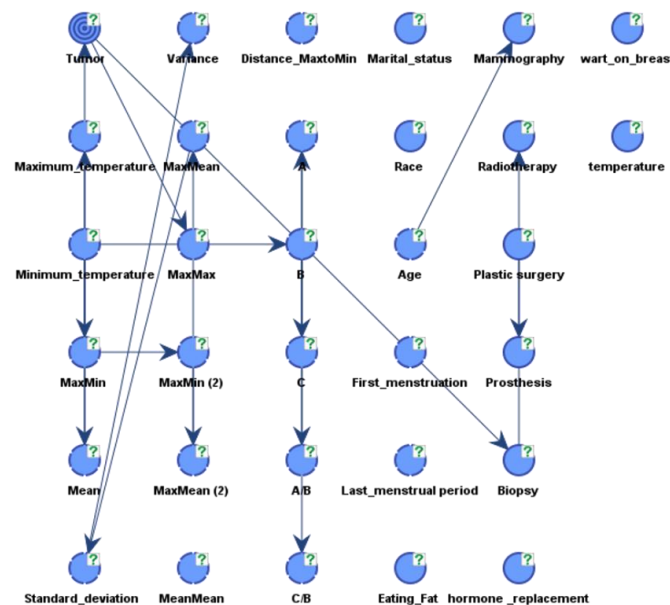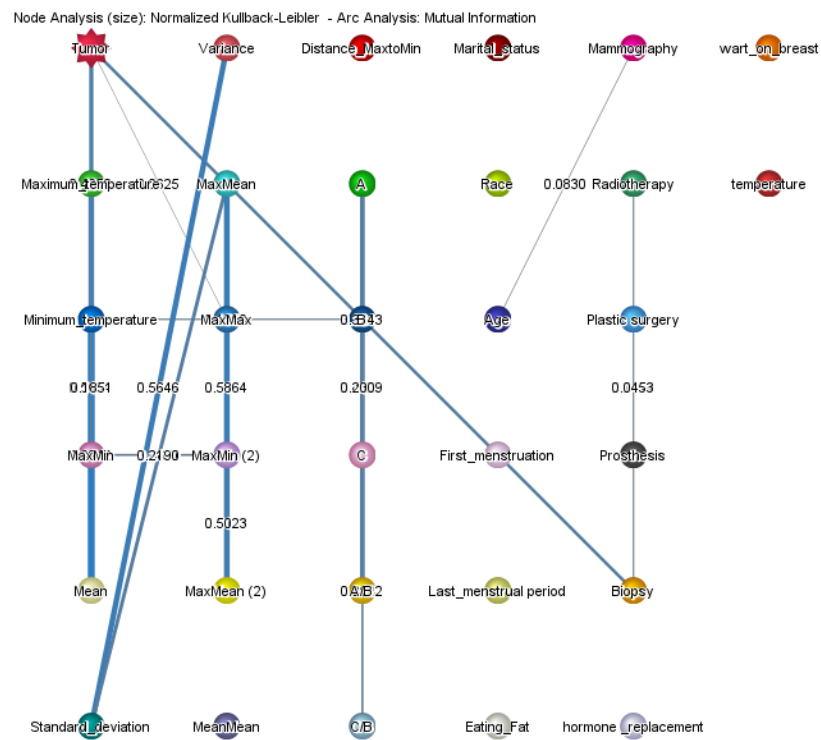

Figure 7. BN with influences from unsupervised learning.
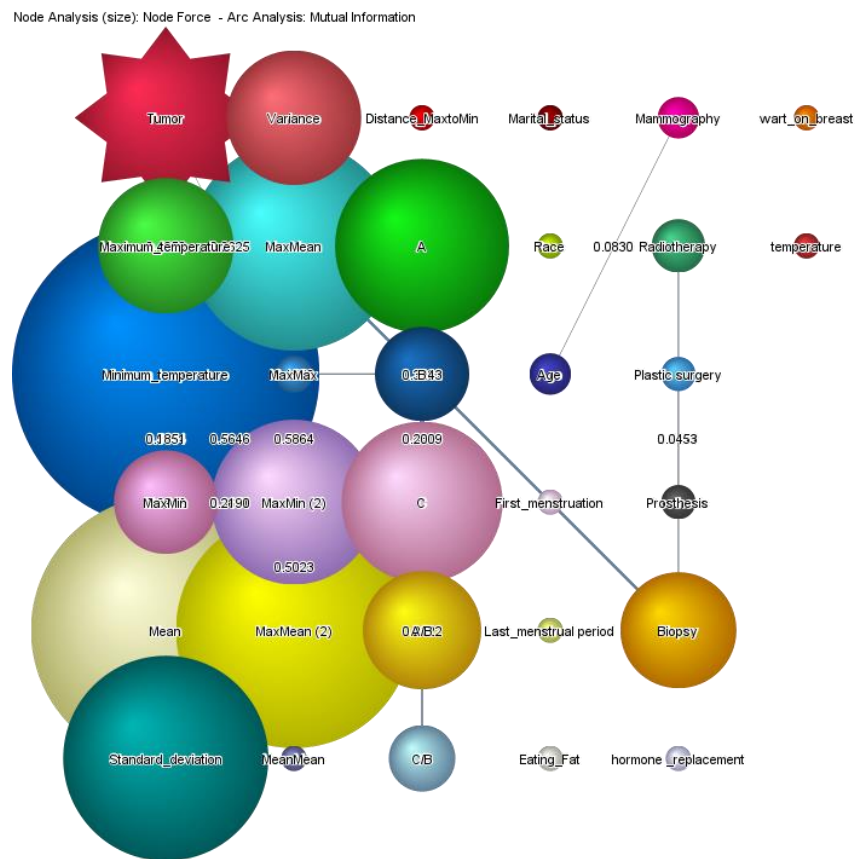
Table 5. Summary of performance indices of BN model.

Target: Tumor	
Value	0
Gini Index	38.2956%
Relative Gini Index	91.1375%
Lift Index	1.5119
Relative Lift Index	97.9255%
ROC Index	95.5709%
Calibration Index	80.8814%
Binary Log-Loss	0.4745

Table 6. Elements show the correctly and incorrectly predicted test data for each class of BN + CNN model.

Occurrences			
Value	0 (178)	1 (128)	Missing Value (1)
0 (179)	164	14	1
1 (128)	14	114	0
Reliability			
Value	0 (178)	1 (128)	Missing Value (1)
0 (179)	91.6201%	7.8212%	0.5587%
1 (128)	10.9375%	89.0625%	0.0000%
Precision			
Value	0 (178)	1 (128)	Missing Value (1)
0 (179)	92.1348%	10.9375%	100.0000%
1 (128)	7.8652%	89.0625%	0.0000%

Figure 8. BN structure from unsupervised learning with node strength evaluated with normalized KL divergence and Arc strength evaluated with mutual information.

Figure 9. BN structure from unsupervised learning with node strength evaluated with "node force" and Arc strength evaluated with mutual information.
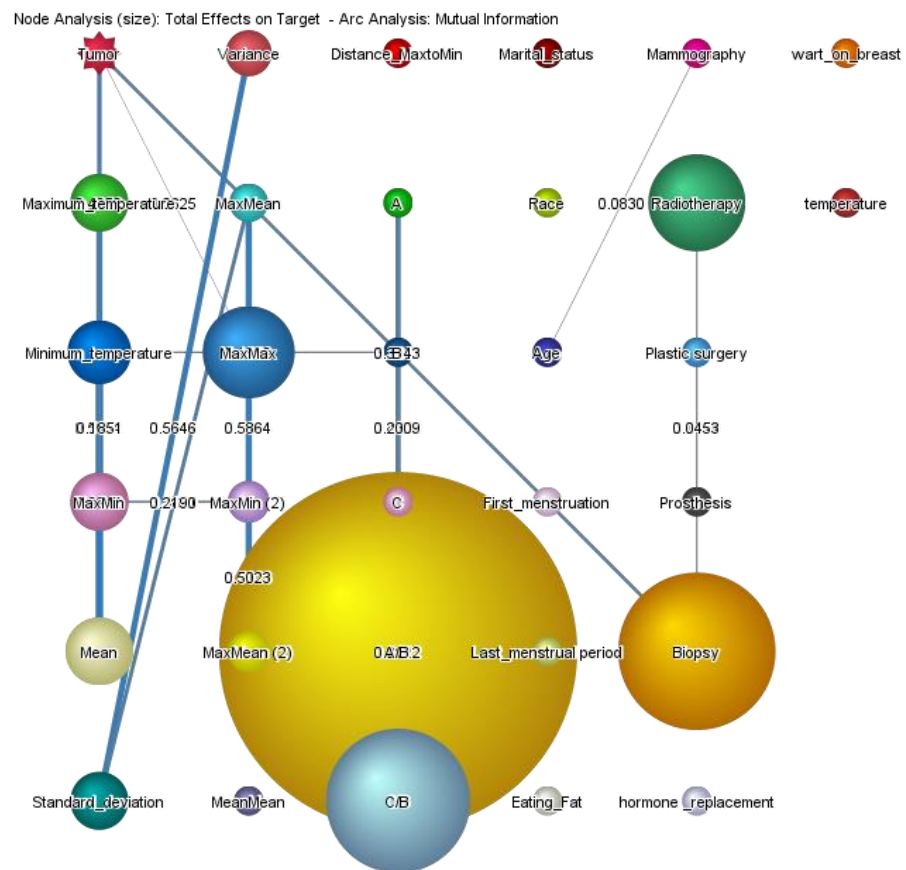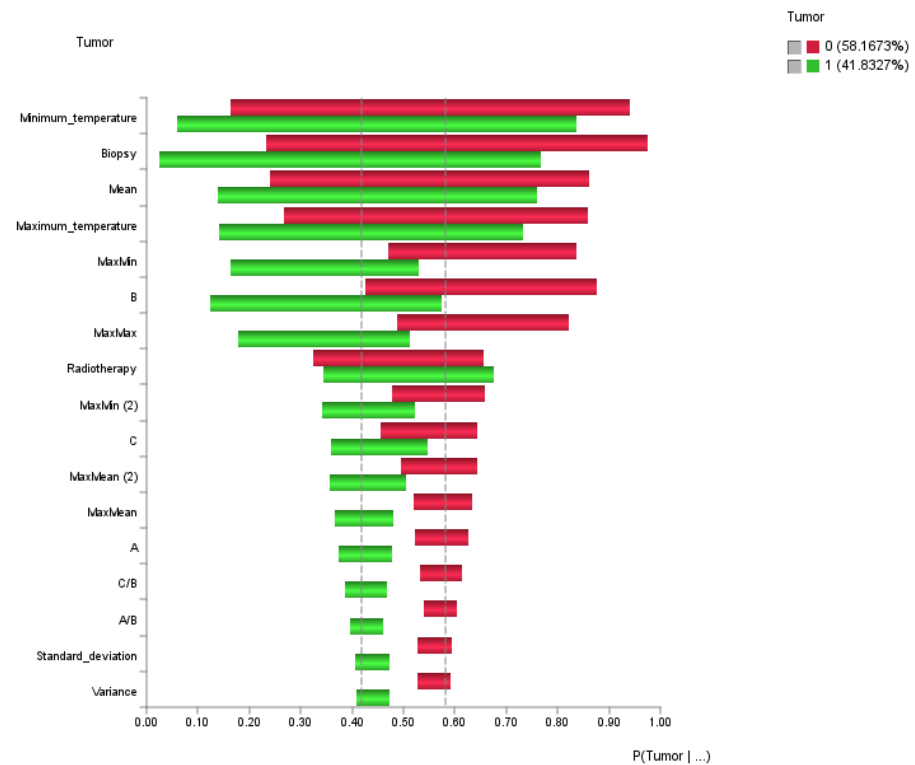
Figure 10. Total effects for the target node tumor.

Figure 11. Tornado effects (conditional probabilities) for the target variable "tumor".
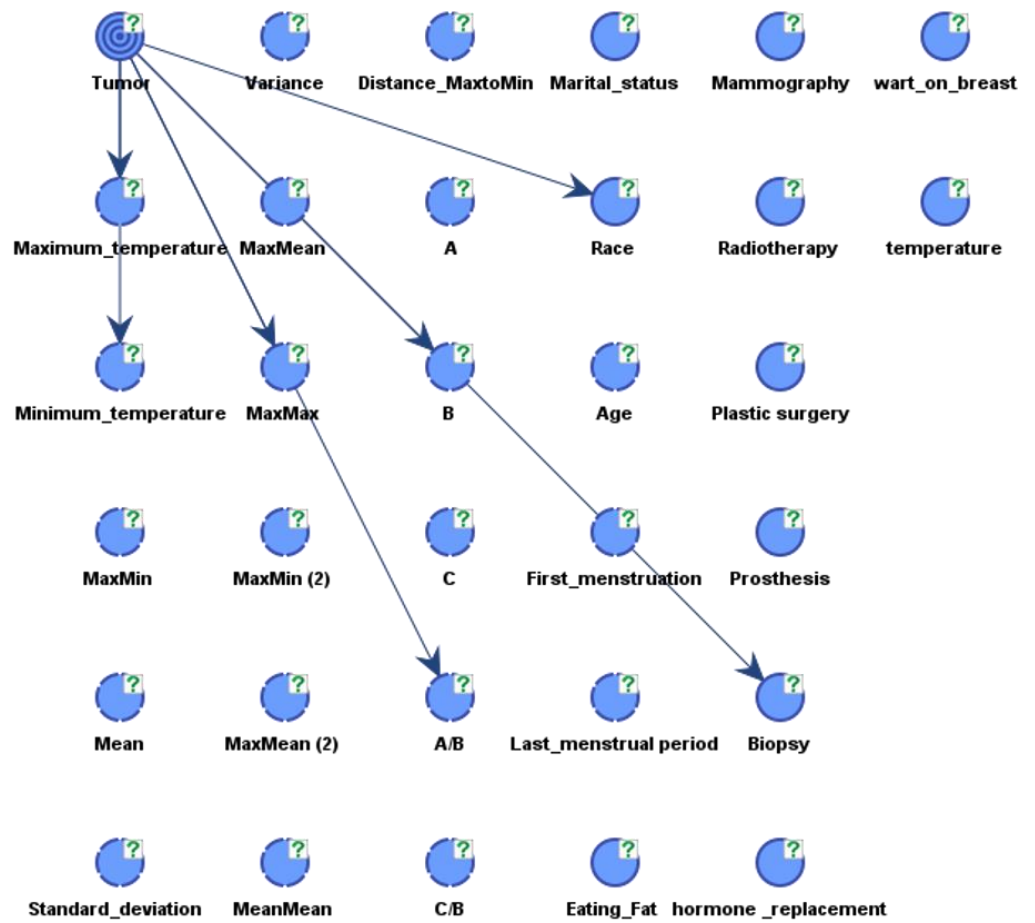
Figure 12. BN influences model for the target variable with supervised learning.
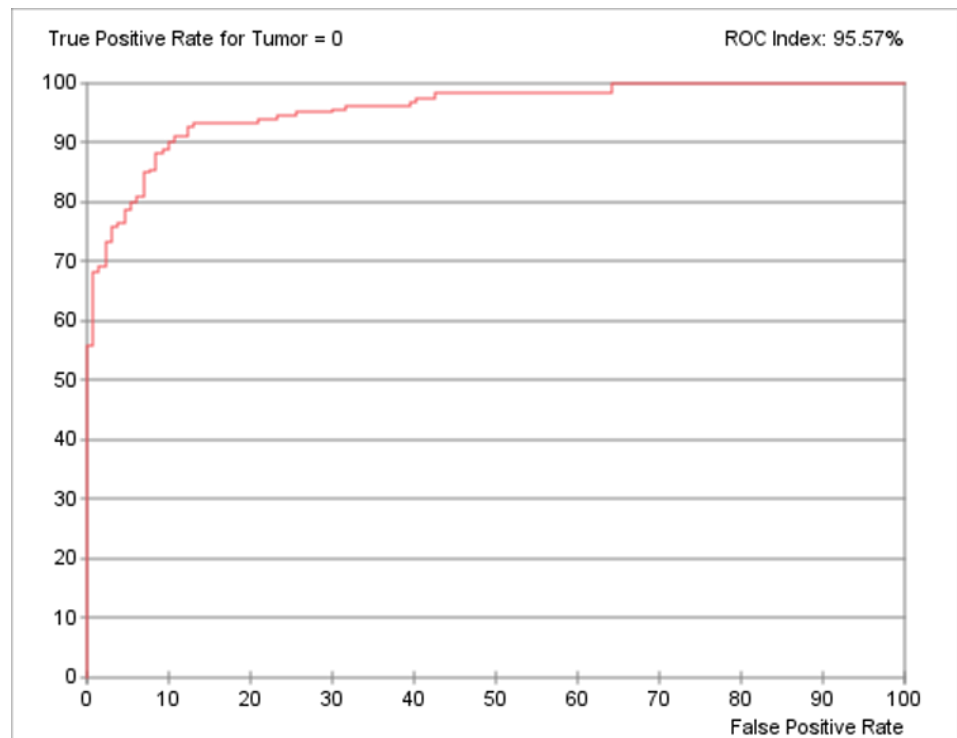

Figure 13. ROC curve for the expert model.

Both expert models were also evaluated with the K-fold validation method. No weaknesses were detected. The results show that both BN models are competitive with the pure CNN diagnostic tool. This means that the BN-constructed expert models can be safely used to understand which factors play a key role in the final successful diagnosis of whether a patient has a tumor (something that is not possible with the pure CNN system).

4. Conclusions and Discussion

An interpretable integrated diagnostic approach has been developed and demonstrated to be effective and accurate even with limited datasets. A number of transfer learning models are also investigated and compared. This study demonstrates the feasibility of developing a computer-aided diagnostic system that can assist doctors in reliably and quickly distinguishing thermograms with or without tumors. However, the model's performance is complicated by the sparse data. More breast thermal images from patients with cancerous cells and individuals without cancerous cells will help the model become more reliable and accurate. Overall, the usage of the transfer learning approach yields better results compared to the baseline approach. One of the best results, by most of the metrics results, is obtained by MobileNet, at 93.8% accuracy amongst the four transfer learning models. The application of transformer models for this study also shows competitive results even compared with the state-of-the-art transfer learning models. These approaches demonstrated high performance compared to the previously published works, such as in [11,12].

Finally, this work reports the successful construction of expert models with the help of BNs and a combination of BNs and CNNs. Two BN-based expert systems have been designed, one using image features and medical data and the second using the previous data plus a CNN prediction. The analysis showed that all these expert models have similarly high accuracy, with BN + CNN models at 90.85% and ResNet50 achieving 93.8%. Thus, it can be claimed and concluded that the crucial factors which drive the decision in the BN models comprise a trusted interpretable diagnosis model which is useful both for a patient and a physician. The developed integrated system can be easily implemented on low-cost portable devices for automatic mass screening and breast self-examination (BSE) as envisaged by the WHO to eradicate breast cancer.

Author Contributions: Conceptualization, N.A. and E.Y.K.N.; Methodology, Y.Z., V.Z. and E.Y.K.N.; Software, N.A. and V.Z.; Validation, Y.Z.; Formal analysis, N.A. and V.Z.; Resources, A.M. (Aigerim Mashekova), A.M. (Anna Midlenko) and O.M.; Data curation, A.M. (Anna Midlenko); Writing—original draft, N.A. and A.M. (Aigerim Mashekova); Writing—review & editing, Y.Z., E.Y.K.N. and V.Z.; Visualization, V.Z. and O.M.; Supervision, Y.Z.; Project administration, A.M. (Aigerim Mashekova), Y.Z. and O.M.; Funding acquisition, A.M. (Aigerim Mashekova); V.Z. for the Bayesian networks design and construction. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The presented research was certified by the Institutional Research Ethics Committee (IREC) of Nazarbayev University (identification number: 592/21072022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The first set of data that support the findings of this study are openly available in Visual Lab at http://visual.ic.uff.br/dmi/, reference number [11]. The second set of data were generated at Astana Multifunctional Hospital by mammologist, oncologist. Derived data supporting the findings of this study are available from the corresponding author Michael Yong Zhao on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Francis, S.V.; Sasikala, M.; Jaipurkar, S.D. *Detection of Breast Abnormality Using Rotational Thermography*; Springer: Singapore, 2017; pp. 133–158.
2. Singh, D.; Singh, A.K. Role of image thermography in early breast cancer detection—Past, present and future. *Comput. Methods Programs Biomed.* **2020**, *183*, 105074. [CrossRef]
3. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
4. World Health Organization (WHO). WHO—Breast Cancer: Prevention and Control. 2020. Available online: http://www.who.int/cancer/detection/breastcancer/en/index1.html (accessed on 5 March 2022).
5. Mashekova, A.; Zhao, Y.; Ng, E.Y.; Zarikas, V.; Fok, S.C.; Mukhmetov, O. Early detection of the breast cancer using infrared technology—A Comprehensive Review. *Therm. Sci. Eng. Prog.* **2022**, *27*, 101142. [CrossRef]
6. Kandlikar, S.G.; Perez-Raya, I.; Raghupathi, P.A.; Gonzalez-Hernandez, J.L.; Dabydeen, D.; Medeiros, L.; Phatak, P. Infrared imaging technology for breast cancer detection—Current status, protocols and new directions. *Int. J. Heat Mass Transf.* **2017**, *108*, 2303–2320. [CrossRef]
7. Kamal, R.; Mansour, S.; Farouk, A.; Hanafy, M.; Elhatw, A.; Goma, M.M. Contrast-enhanced mammography in comparison with dynamic contrast-enhanced MRI: Which modality is appropriate for whom? *Egypt. J. Radiol. Nucl. Med.* **2021**, *52*, 216. [CrossRef]
8. Khan, S.; Rahmani, H.; Shah, S.A.A.; Bennamoun, M. *A Guide to Convolutional Neural Networks for Computer Vision*; Morgan Claypool Publishers: San Rafael, CA, USA, 2018.
9. Sarigoz, T.; Ertan, T. Role of dynamic thermography in diagnosis of nodal involvement in patients with breast cancer: A pilot study. *Infrared Phys. Technol.* **2020**, *108*, 103336. [CrossRef]
10. Saniei, E.; Setayeshi, S.; Akbari, M.E.; Navid, M. Parameter estimation of breast tumor using dynamic neural network from thermal pattern. *J. Adv. Res.* **2016**, *7*, 1045–1055. [CrossRef] [PubMed]
11. Torres-Galván, J.C.; Guevara, E.; González, F.J. Comparison of deep learning architectures for pre-screening of breast cancer thermograms. In Proceedings of the 2019 Photonics North (PN), Quebec City, QC, Canada, 21–23 May 2019; pp. 1–2.
12. Zuluaga-Gomez, J.; Al Masry, Z.; Benaggoune, K.; Meraghni, S.; Zerhouni, N. A CNN-based methodology for breast cancer diagnosis using thermal images. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2021**, *9*, 131–145. [CrossRef]
13. Bapin, Y.; Zarikas, V. Smart Building's Elevator with Intelligent Control Algorithm based on Bayesian Networks. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 16–24. [CrossRef]
14. Zarikas, V. Modeling decisions under uncertainty in adaptive user interfaces. *Univers. Access Inf. Soc.* **2007**, *6*, 87–101. [CrossRef]
15. Zarikas, V.; Papageorgiou, E.; Regner, P. Bayesian network construction using a fuzzy rule based approach for medical decision support. *Expert Syst.* **2015**, *32*, 344–369. [CrossRef]
16. Amrin, A.; Zarikas, V.; Spitas, C. Reliability analysis and functional design using Bayesian networks generated automatically by an "Idea algebra" framework. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 211–225. [CrossRef]
17. Amrin, A.; Zarikas, V.; Spitas, C. Reliability analysis of an automobile system using idea algebra method equipped with dynamic Bayesian network. *Int. J. Reliab. Qual. Saf. Eng.* **2022**, *29*, 2150045. [CrossRef]
18. Darmeshov, B.; Zarikas, V. Efficient Bayesian expert models for fever in neutropenia and fever in neutropenia with bacteremia. In *Proceedings of the Future Technologies Conference*; Springer: Cham, Switzerland, 2019; pp. 124–143. [CrossRef]
19. Hamilton, P.W.; Montironi, R.; Abmayr, W.; Bibbo, M.; Anderson, N.; Thompson, D.; Bartels, P.H. Clinical applications of Bayesian belief networks in pathology. *Pathologica* **1995**, *87*, 237–245. [PubMed]
20. Liu, S.; Deng, W. Very deep convolutional neural network based image classification using small training sample size. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 730–734. [CrossRef]
21. Benbrahim, H.; Behloul, A. Fine-tuned Xception for Image Classification on Tiny ImageNet. In Proceedings of the 2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP), El Oued, Algeria, 20–21 November 2021; pp. 1–4. [CrossRef]
22. Aidossov, N.; Mashekova, A.; Zhao, Y.; Zarikas, V.; Eddie-Yin-Kwee Ng Mukhmetov, O. Intelligent Diagnosis of Breast Cancer with Thermograms using Convolutional Neural Networks. In Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022), Virtual, 3–5 February 2022; Volume 2, pp. 598–604. [CrossRef]
23. Elkan, C. The foundations of cost-sensitive learning. In *Proceedings of the 17th international Joint Conference on Artificial Intelligence—Volume 2 (IJCAI'01)*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2001; pp. 973–978.
24. Udas, N.; Beuth, F.; Kowerko, D. Concept Detection in Medical Images using Xception Models—TUC_MC at ImageCLEFmed. In Proceedings of the CLEF 2020, (Working Notes), Thessaloniki, Greece, 22–25 September 2020.
25. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
26. Silva, L.F.; Saade, D.C.M.; Sequeiros, G.O.; Silva, A.C.; Paiva, A.C.; Bravo, R.S.; Conci, A. A new database for breast research with infrared images. *J. Med. Imaging Health Inform.* **2014**, *4*, 92–100. [CrossRef]
27. Visual Lab DMR Database. Available online: http://visual.ic.uff.br/dmi/ (accessed on 26 November 2022).

28.	Andersson, E.; Berglund, R. *Evaluation of Data Augmentation of MR Images for Deep Learning*; Lund University: Lund, Sweden, 2018.
29.	Dong, Q.; Gong, S.; Zhu, X. Imbalanced Deep Learning by Minority Class Incremental Rectification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1367–1381. [CrossRef] [PubMed]