Journal of Intelligent Learning Systems and Applications, 2024, 16, 262-273
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

A Hybrid Learning Algorithm for Breast Cancer Diagnosis

Alio Boubacar Goga*, Harouna Naroua, Chaibou Kadri

Laboratoire d'Informatique Fondamentale et Appliquée-Sciences de l'Ingénierie (LIFA-SI), Département de Mathématiques et Informatique, Faculté des Sciences et Techniques, Université Abdou Moumouni, Niamey, Niger
Email: *gogaalioboubacar@gmail.com

How to cite this paper: Goga, A.B., Naroua, H. and Kadri, C. (2024) A Hybrid Learning Algorithm for Breast Cancer Diagnosis. *Journal of Intelligent Learning Systems and Applications*, **16**, 262-273.
https://doi.org/10.4236/jilsa.2024.163014

Received: July 20, 2024
Accepted: August 26, 2024
Published: August 29, 2024

Abstract

In many fields, particularly that of health, the diagnosis of diseases is a very difficult task to carry out. Therefore, early detection of diseases using artificial intelligence tools can be of paramount importance in the medical field. In this study, we proposed an intelligent system capable of performing diagnoses for radiologists. The support system is designed to evaluate mammographic images, thereby classifying normal and abnormal patients. The proposed method (DiagBC for Breast Cancer Diagnosis) combines two (2) intelligent unsupervised learning algorithms (the C-Means clustering algorithm and the Gaussian Mixture Model) for the segmentation of medical images and an algorithm for supervised learning (a modified DenseNet) for the diagnosis of breast images. Ultimately, a prototype of the proposed system was implemented for the Magori Polyclinic in Niamey (Niger) making it possible to diagnose (or classify) breast cancer into two (2) classes: the normal class and the abnormal class.

Keywords

Image Diagnosis, Segmentation, DenseNet, Unsupervised Learning, Supervised Learning, Breast Cancer

1. Introduction

Breast cancer is one of the major cancers affecting women's lives. However, the standard procedure for diagnosing it by pathologists is time-consuming [1]. Therefore, the development of an intelligent system is a way to assist pathologists during diagnosis. Benign tumors (considered normal) grow slowly [2]. So, these types of tumors are harmless.

In traditional Machine Learning (ML) algorithms, features are used to diagnose images. These methods consist of extracting the characteristics of each image from the dataset manually and then training a classifier on these characteristics. These supervised learning methods can provide very good results. However, their performance strongly depends on the quality of the characteristics previously found.

Transfer learning is a better approach that uses a pre-trained model to extract unique features from each medical image. Additionally, the current performance of these models caught our attention as they achieved better results in feature extraction and classification of breast cancer, especially the DenseNet [3].

Image segmentation algorithms, particularly mammography and histopathology images, are currently used in the field of computer vision and disease diagnosis. They permit good and early (or rapid) diagnosis of breast cancer. The fuzzy c-means method is an unsupervised clustering method (unsupervised learning) and is one of the most used methods for medical image segmentation.

Breast cancer is the second leading cause of death among women. In addition, there is a shortage of specialist doctors in our country, perhaps even worldwide. This global scourge is very difficult to diagnose at an early stage. Therefore, the creation of Intelligent Systems (IS) is necessary in our hospitals and clinics to help radiologists to properly diagnose breast cancer and other cancers.

In our hospitals, doctors use several techniques to diagnose breast cancer including mammography, ultrasound and Magnetic Resonance Imaging (MRI). Currently, mammography and ultrasound are commonly used diagnostic techniques [4].

Mammography is ineffective in patients under 40, with dense breasts and less sensitive to small tumors. Most breast diseases resemble signs of cancer and require additional tests to make a correct diagnosis (these tests are very expensive for most patients). To overcome these problems, we proposed an intelligent system to help radiologists make a final decision.

Nowadays, mammographic images exist in abundance in our hospitals and clinics. This is caused by the number of mammograms performed per woman. However, these images are not processed in most cases. These images obtained will be used by statistical (or automatic) learning algorithms with the aim of to obtain an intelligent or artificial intelligence model. In our hospitals, there are two (2) types of data: annotated or labeled data and unlabeled data. We used unsupervised learning algorithms to solve the lack of annotated data at Polyclinic Magori.

In this study, we propose a hybrid learning algorithm for breast cancer diagnosis. It combines unsupervised and supervised learning and the data used was collected from Magori Polyclinic in Niamey (Niger). The proposed system uses intelligent algorithms (K-means clustering, Fuzzy C-means, and modified DenseNet) based on our own experience [4] and the literature review regarding this study.

2. Related Works

In [5] an intelligent model based on modified DenseNet was proposed to diagnose histopathology images of breast cancer disease into two (2) classes: the benign class and the malignant class.

In [6], it was realized that manual segmentation of medical images is time-consuming and does not take into consideration the appearance of anatomical structures. Therefore, to solve this problem, the authors proposed a clinical target volume (CTV) auto-segmentation method called Deep Dilated Residual Network (DD-ResNet). The results show that the self-segmented contours of the CTV were close to the manually segmented contours. The deep learning algorithm (DD-ResNet) proposed by the authors could be used to improve consistency in the contouring and streamlining of breast cancer radiotherapy processes.

In [7], to effectively help radiologists in making accurate diagnoses, an intelligent system was proposed for breast cancer screening. This system requires detection, segmentation, and classification of the disease using deep learning methods. To recognize the mass and classify it as benign or malignant, a deep convolutional neural network (CNN) is applied. To evaluate the proposed system, the INbreast database was used and according to the authors, the proposed system could be used to help radiologists in all stages of detection, segmentation, and classification of breast masses. The below **Figure 1** shows the diagram of the intelligent system.

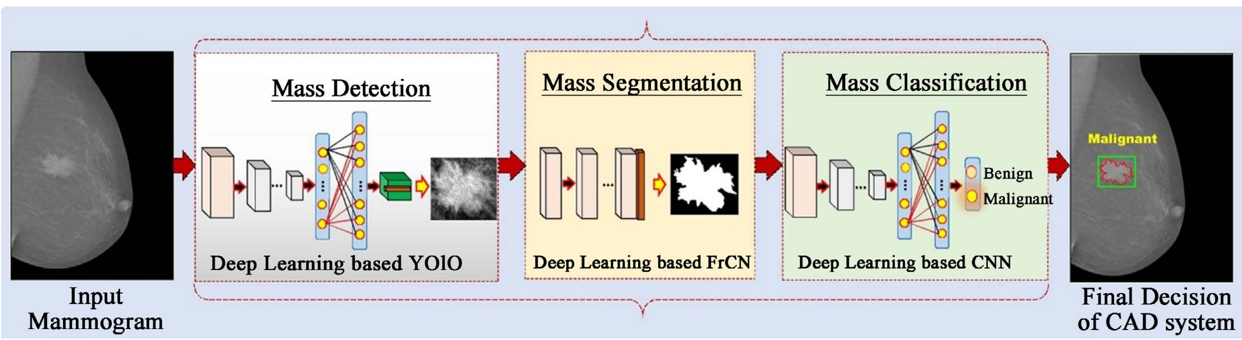

Figure 1. Intelligent System in [7].

In [8], the authors identified seven (7) connected bras. These devices are currently an alternative technique to mammography, given the shortcomings of the previous devices, particularly the detection of breast cancer in dense breasts. Note that these devices are currently in the testing stage. They use technologies like thermal sensors, electrical impedance tomography, and ultrasound. Finally, these connected bras use Artificial Intelligence (AI) methods.

In [9], coronavirus disease (COVID-19) emerged in October 2019. The virus that caused this scourge was called coronavirus 2 severe acute respiratory syndrome (SARS-CoV-2). According to the World Health Organization (WHO), at the end of January, this disease caused more than 2 million deaths. So, an automatic detection system should be introduced as a rapid alternative diagnosis to prevent the spread of COVID-19 among humans. Among AI techniques, these authors chose deep learning including DenseNet, Inception-V3, and Inception-ResNetV4 to detect this disease using X-ray images. They were able to improve DenseNet and also showed that it is more efficient than the two (2) other models cited above. The pre-trained DenseNet improved by the authors achieved 92% accuracy while Inception-V3 and Inception-ResNetV4 achieved accuracies of 83.47% and 85.57%, respectively. Finally, in their study, the authors proposed a deep transfer learning method based on chest X-ray images collected from COVID-19 patients. It also provides insight into how deep transfer learning approaches can be used to detect COVID-19 at an early stage. Figure 2 shows the basic architecture of DenseNet models.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 24	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48	[1 × 1 conv, 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 16	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 2. Basic architecture of DenseNet models in [10].

In [10], a potential new approach combining DenseNet201 and XGBoost was proposed. This type of system can effectively identify the presence of cancer cells. To experiment with the proposed DLXGB method (for Deep Learning and eXtreme Gradient Boosting), the authors used a dataset of histopathological images of breast cancer. The choice of this database is based on its use by many researchers for the creation of intelligent systems particularly in [11] [12].

In [13], the robust properties of CNN (Convolutional Neural Network) and its major success in solving various computer vision problem tasks make it a plausible candidate among classifiers concerning the classification of images and object recognition.

Data partitioning is a method of analyzing medical images that consists of separating the n objects of a database into k subsets so that each group (or subset) has similar objects and that the groups are well different from each other [14]. It makes it possible to detect hidden structures in datasets without prior knowledge. Several different approaches exist; the methods are distinguished by the nature of the partitions created.

Figure 3 shows the segmentation method of medical images (Normal Image, Benign Image and Malignant Image) proposed in [15]:

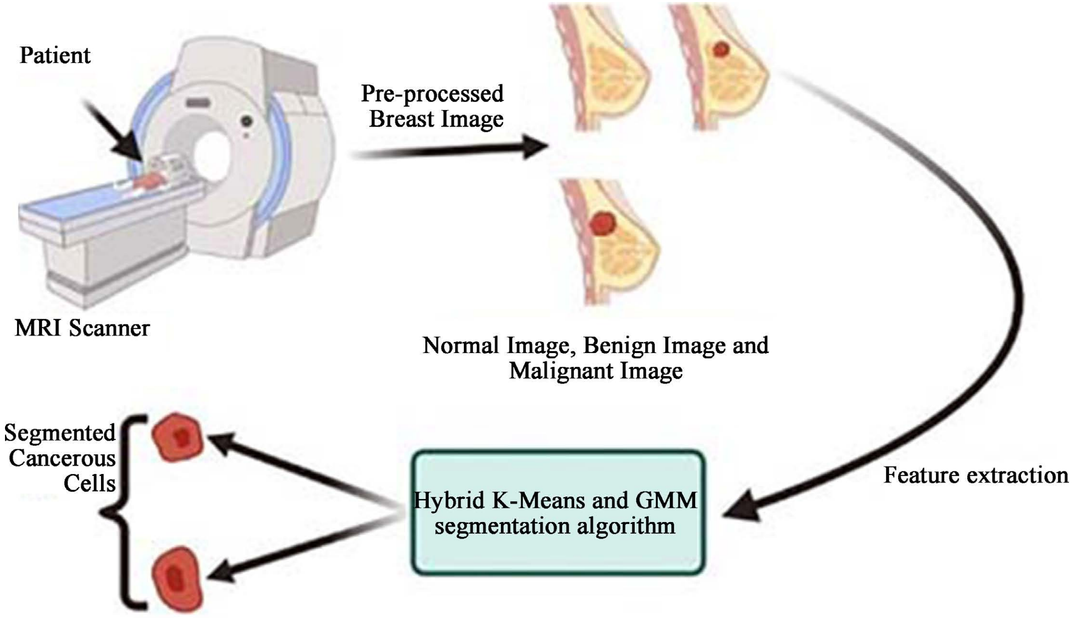

Figure 3. Proposed model for breast cancer segmentation in [15].

3. Materials and Methods

We were able to collect 80 medical images of the breast (mammographic images) at the Magori Polyclinic in Niamey. These images were acquired in DCM or DICOM (for Digital Imaging Communication in Medicine) format and were converted to PNG (Portable Network Graphics) while maintaining interpretable image quality.

Figure 4 shows the functional diagram of the proposed method.

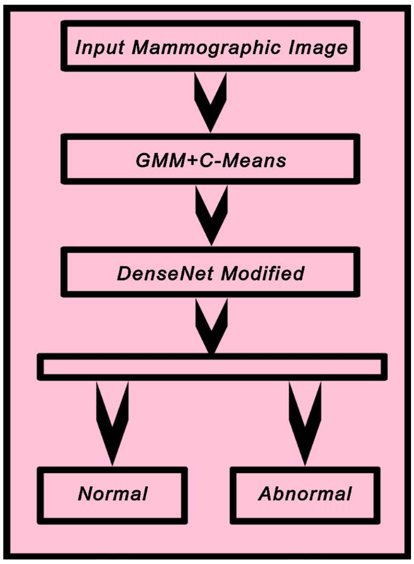

Figure 4. Proposed approach for the diagnosis of breast cancer (DiagBC).

This work shows a combination of supervised and unsupervised learning techniques. We were able to combine two unsupervised learning methods namely C-Means and the Gaussian Mixture Model (GMM) for the segmentation of mammographic images on one hand and on the other hand a supervised learning technique (a modified DenseNet), for the diagnosis of breast cancer. Therefore, the proposed method (DiagBC for Diagnosis of Breast Cancer) is a combination of supervised and unsupervised learning, making it possible to classify this disease into two categories: normal and abnormal.

We modified the DenseNet by adding a dense layer of 512 perceptrons.

4. Results and Discussion

In this part, we will discuss our test results in order to reach to the appropriate method. We started collecting mammographic images at the Magori Polyclinic in Niamey aiming at helping radiologists in decisions making. Figure 5 shows the first test performed on the unlabeled mammogram images:

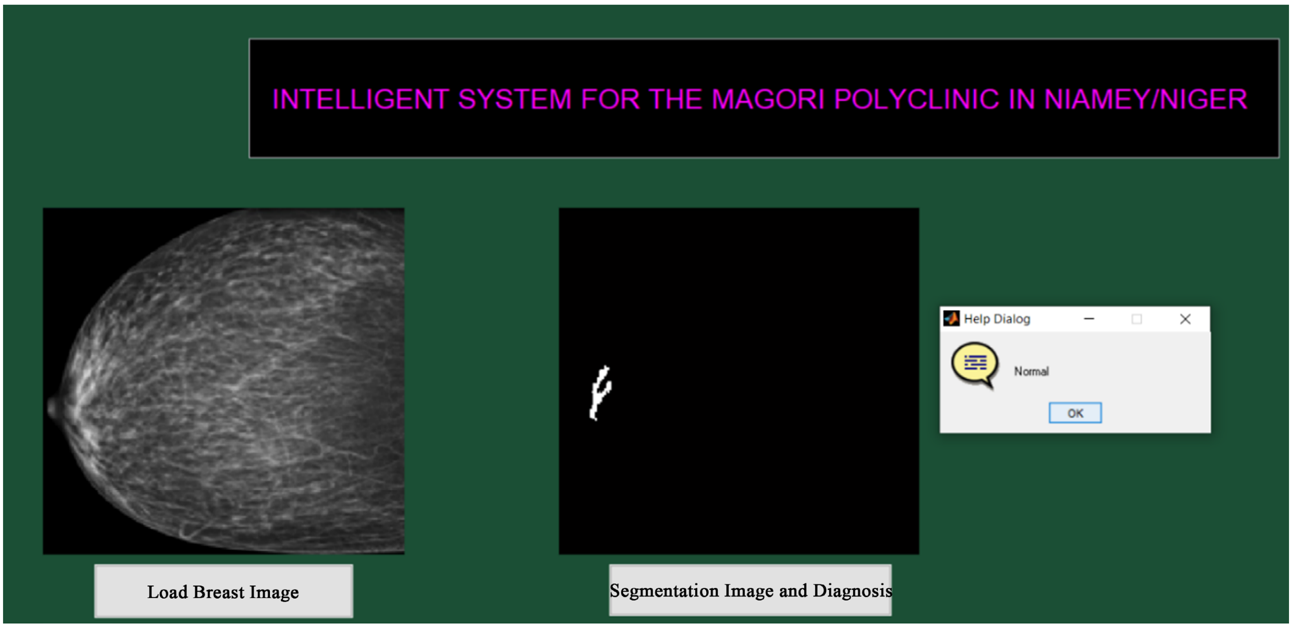

Figure 5. Image segmentation and diagnosis (Test 1).

Figure 5 shows the segmentation of a mammographic image using C-Means and GMM and the diagnosis of breast cancer using the modified DenseNet method. The result shows a healthy patient.

Figure 6 shows the second test performed on the unlabeled mammogram images:

This patient is also normal according to the proposed system. Figure 7 shows another diagnosis carried out.

Figure 8 shows the segmentation of a mammographic image using C-Means and GMM and the diagnosis of breast cancer using the modified DenseNet method. The system shows that the patient is abnormal.

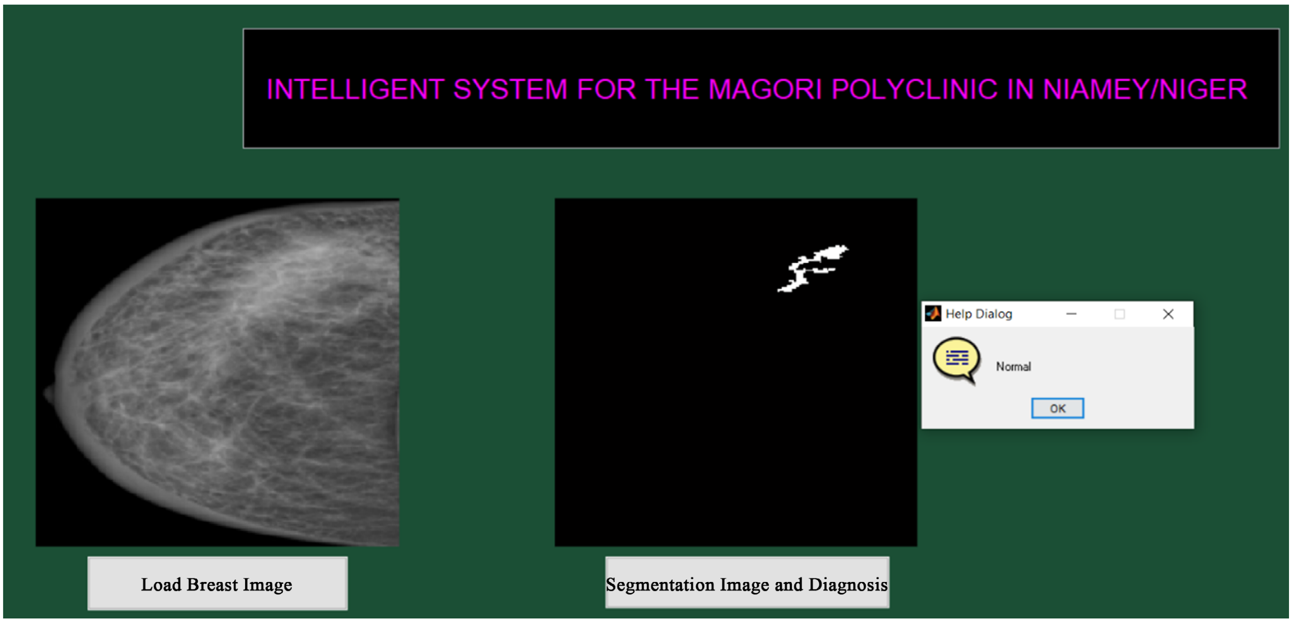

Figure 6. Image segmentation and diagnosis (Test 2).

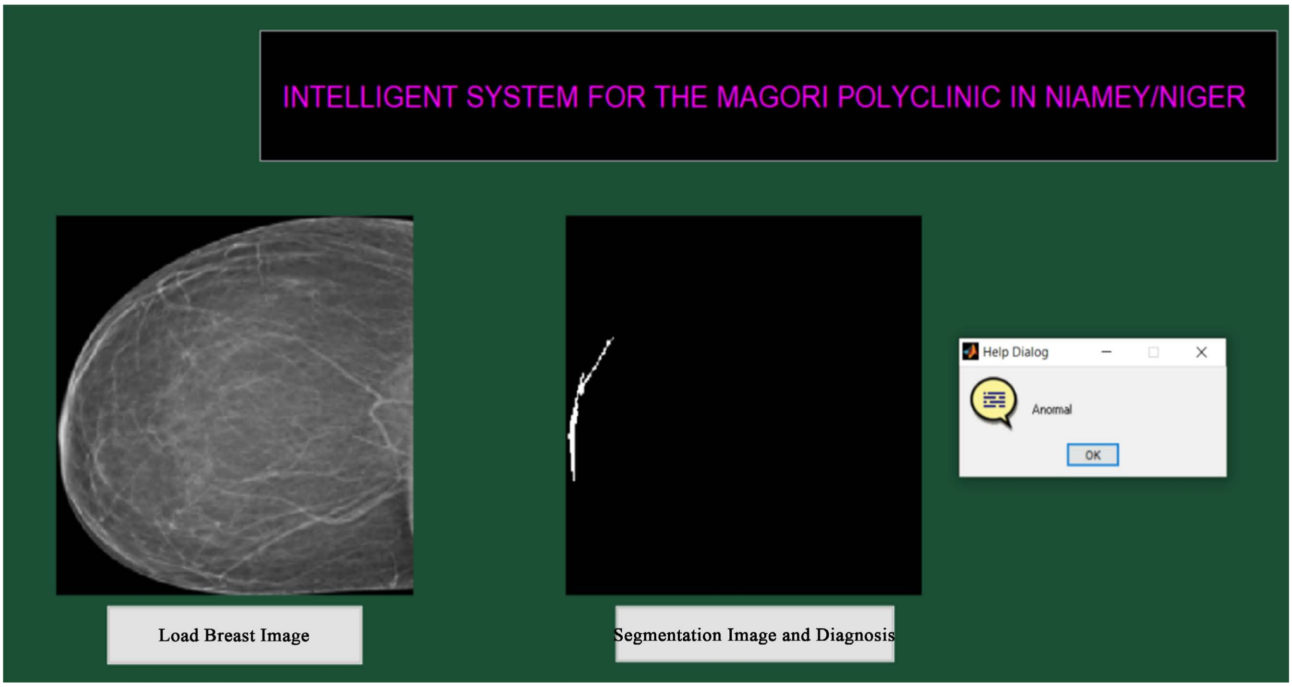

Figure 7. Image segmentation and diagnosis (Test 3).

To verify that the system has good generalization, we found it useful to test it on another database. A concise description of this database is hereby given.

The diagnostic Breast Cancer Histopathology Images (BreakHis) is composed of 7909 microscopic images of breast lesion tissues collected from 82 patients using different magnification factors (40×, 100×, 200×, 400×). It contains 2480 benign samples and 5429 malignant samples. This database was built in collaboration with the P&D-Anatomy Pathological and Cytopathology laboratory, in

Parana, Brazil [16].

To check whether the system gives a good result, we tested it on an abnormal mammogram from the BreakHis database.

This patient is abnormal according to the system. Figure 9 shows the normal mammogram test from P&D Laboratory, Parana, Brazil.

The proposed system gives the same result as the robotic vision and imaging

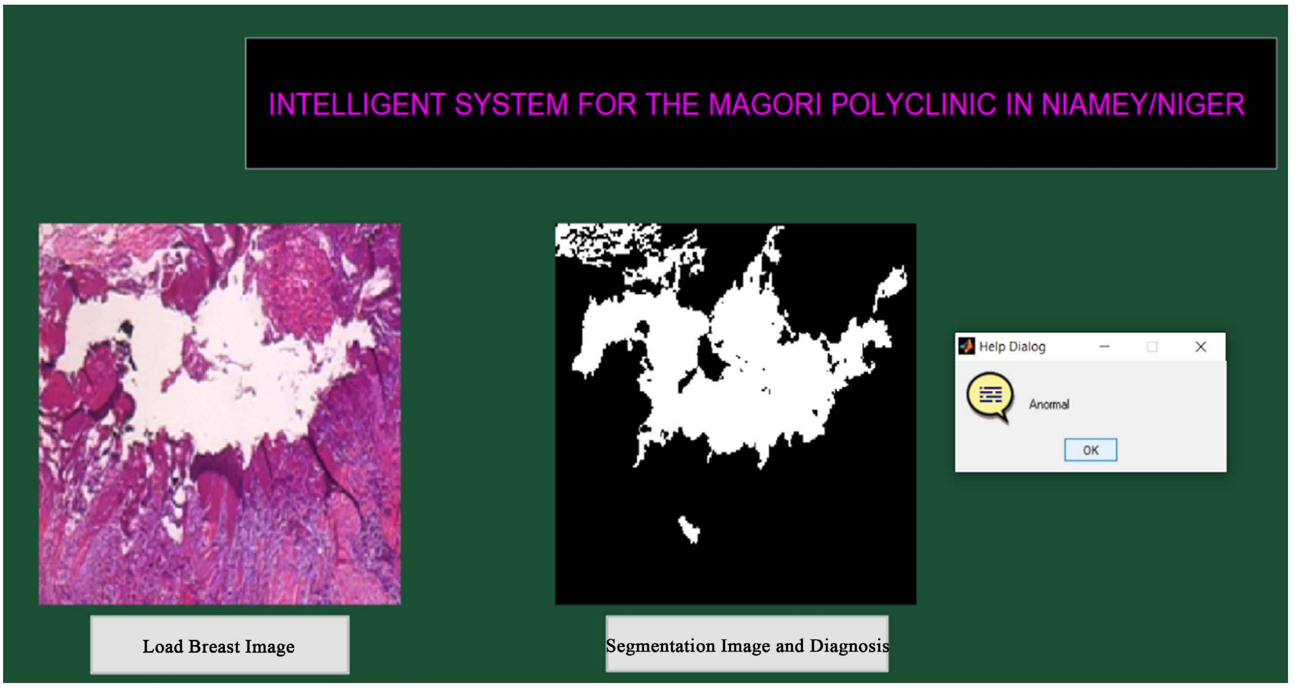

Figure 8. Image segmentation and diagnosis (Test 4).

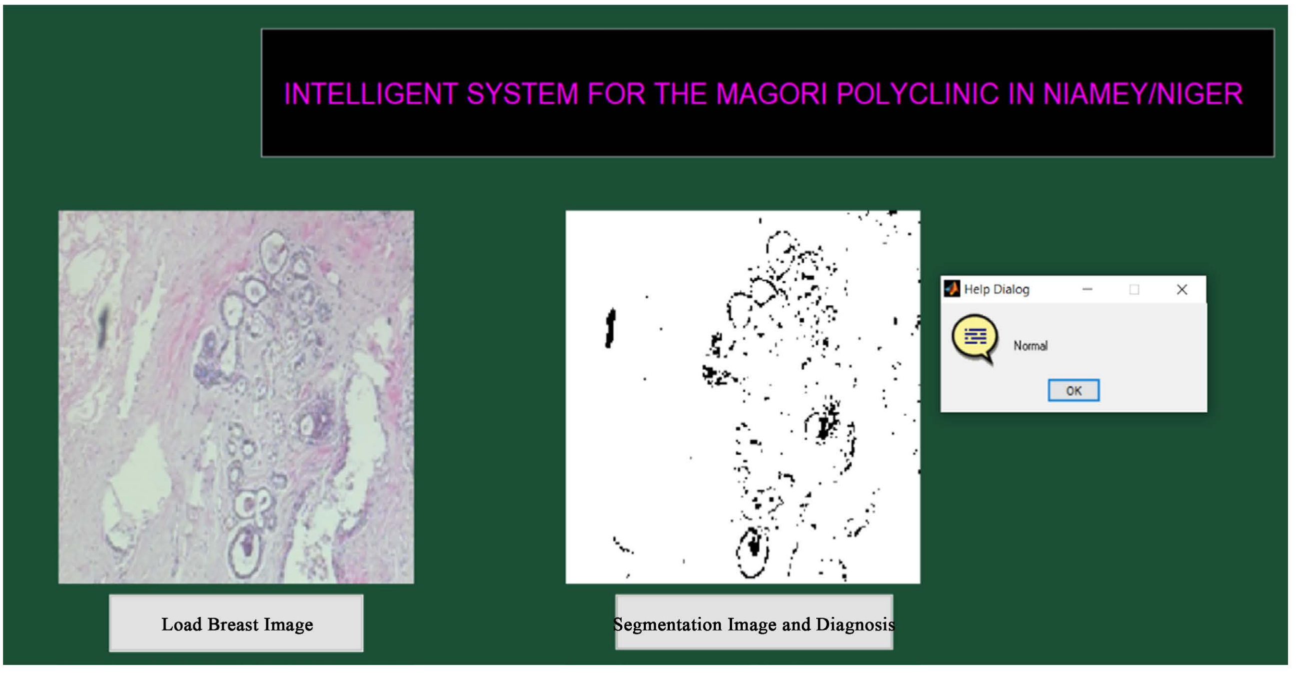

Figure 9. Image segmentation and diagnosis (Test 5).

laboratory, that is to say, that the patient is normal. The Accuracy (acc on Figure 10) increases with the number of epochs (x axis on Figure 10). Figure 10 shows the Accuracy and Epoch of the new model while Figure 11 shows the Loss.

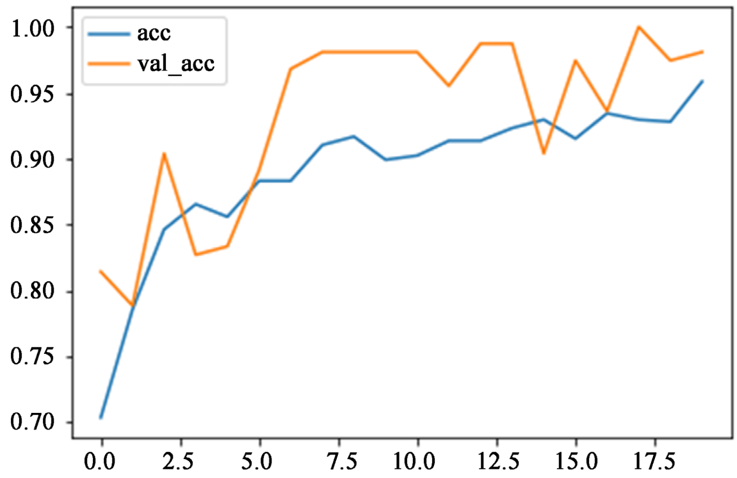

Figure 10. Accuracy.

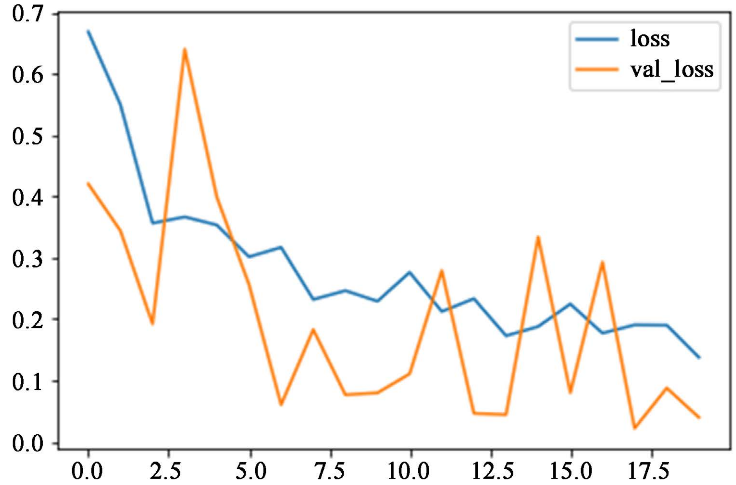

Figure 11. Loss.

Figure 10 and Figure 11 show that the proposed hybrid model has good generalization. Moreover, the performance metrics of the model, on test samples, were determined. Table 1 shows these metrics.

Table 1. Performance metrics of DiagBC.

Accuracy	Recall	Precision	F1 Score
99.58%	96.11%	95%	95.55%

With:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \quad (1)$$

$$\text{Recall} = \frac{TN + TP}{TP + FN} \times 100 \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \tag{3}$$

$$F1-\text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{4}$$

The different performance metrics used to evaluate the proposed approach are Accuracy, Recall, Precision, and F1 score as shown in **Table 1**. TN, TP, FP, and FN represent the number of True Negatives, True Positives, False Positives, and False, respectively.

Precision is the ratio between the number of True Positives (TP) and the sum of True Positives (TP) and False Positives (FP).

A Recall of 1 means that all positive samples have been correctly classified. F1-Score is the weighted average of Precision and Recall.

For a given class, a classifier, and a sample, four cases can arise:

- The sample is of this class, and the classifier is not mistaken: it is a true positive (TP).
- The sample is of this class, but the classifier is wrong: it is a false negative (FN).
- The sample is not of this class, but the classifier assigns it anyway: it is a false positive (FP).
- The sample is not of this class, and the classifier does not place it in this class either: it is a true negative (TN).

5. Conclusion and Future Work

A hybrid learning system for breast cancer early detection and diagnosis has been proposed. A prototype has been implemented and tested with patients from the Magori Polyclinic in Niamey (Niger) and the diagnostic Breast Cancer Histopathology Images (BreakHis) dataset. The results of the proposed system were compared very well with those of the robotic vision and imaging laboratory. The proposed hybrid system detects breast cancer and classifies it into normal and abnormal patients. Finally, the proposed model obtained an accuracy of 99.58%.

For future work, we intend to:

- Diagnose breast cancer in several classes.
- Improve the Accuracy of the proposed model.
- Strengthen the evaluation of the effect of the system applied in practical clinics, including the detailed explanation of the accuracy, sensitivity and specificity of the system.
- Using cross-validation or other statistical methods to evaluate the generalization ability and robustness of the system on different data sets.
- Optimize diagrams to show how the system processes medical images and diagnostic results, ensuring that the diagrams convey the main information

succinctly and clearly.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Liew, X.Y., Hameed, N. and Clos, J. (2021) An Investigation of XGBoost-Based Algorithm for Breast Cancer Classification. *Machine Learning with Applications*, **6**, Article 100154. https://doi.org/10.1016/j.mlwa.2021.100154

[2] Al-Masni, M.A., Al-Antari, M.A., Park, J., Gi, G., Kim, T., Rivera, P., *et al.* (2018) Simultaneous Detection and Classification of Breast Masses in Digital Mammograms via a Deep Learning Yolo-Based CAD System. *Computer Methods and Programs in Biomedicine*, **157**, 85-94. https://doi.org/10.1016/j.cmpb.2018.01.017

[3] Li, X., Shen, X., Zhou, Y., Wang, X. and Li, T. (2020) Classification of Breast Cancer Histopathological Images Using Interleaved Densenet with SENet (IDSNet). *PLOS ONE*, **15**, e0232127. https://doi.org/10.1371/journal.pone.0232127

[4] Singh, B.K. (2019) Determining Relevant Biomarkers for Prediction of Breast Cancer Using Anthropometric and Clinical Features: A Comparative Investigation in Machine Learning Paradigm. *Biocybernetics and Biomedical Engineering*, **39**, 393-409.

[5] Goga, A.B., Kadri, C. and Naroua, H, (2023) Deep Learning for the Recognition of Benign and Malignant Breast Lesions in Medical Images. *ARPN Journal of Engineering and Applied Sciences*, **18**, 2151-2157.

[6] Kuo, M., Tao, Z., Xinyan, C., Bo, C., Yu, T., Shulian, W., Yexiong, L. and Jianrong, D. (2018) Fully Automatic and Robust Segmentation of the Clinical Target Volume for Radiotherapy of Breast Cancer Using Big Data and Deep Learning. *Physica Medica*, **18**, 13-19. https://doi.org/10.1016/j.ejmp.2018.05.006

[7] Al-antari, M.A., Al-masni, M.A., Choi, M., Han, S. and Kim, T. (2018) A Fully Integrated Computer-Aided Diagnosis System for Digital X-Ray Mammograms via Deep Learning Detection, Segmentation, and Classification. *International Journal of Medical Informatics*, **117**, 44-54. https://doi.org/10.1016/j.ijmedinf.2018.06.003

[8] Masry, Z.A., Zerhouni, N., Gay, C., Meraghni, S., Lodi, M., Mathelin, C. and Devalland, C. (2021) Détection du cancer du sein à l'aide de soutiens-gorge connectés en 2021: Analyses et perspectives.

[9] Albahli, S., Ayub, N. and Shiraz, M. (2021) Coronavirus Disease (COVID-19) Detection Using X-Ray Images and Enhanced Densenet. *Applied Soft Computing*, **110**, Article 107645. https://doi.org/10.1016/j.asoc.2021.107645

[10] Alkassar, S., Jebur, B.A., Abdullah, M.A.M., Al-Khalidy, J.H. and Chambers, J.A. (2021) Going Deeper: Magnification-Invariant Approach for Breast Cancer Classification Using Histopathological Images. *IET Computer Vision*, **15**, 151-164. https://doi.org/10.1049/cvi2.12021

[11] Alom, M.Z., Yakopcic, C., Nasrin, M.S., Taha, T.M. and Asari, V.K. (2019) Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *Journal of Digital Imaging*, **32**, 605-617. https://doi.org/10.1007/s10278-019-00182-7

[12] Murtaza, G., Shuib, L., Mujtaba, G. and Raza, G. (2019) Breast Cancer Multi-Classification through Deep Neural Network and Hierarchical Classification Ap-

proach. *Multimedia Tools and Applications*, **79**, 15481-15511. https://doi.org/10.1007/s11042-019-7525-4

[13] Mateen, M., Wen, J., Nasrullah Song, S. and Huang, Z. (2018) Fundus Image Classification Using VGG-19 Architecture with PCA and SVD. *Symmetry*, **11**, Article 1. https://doi.org/10.3390/sym11010001

[14] Jain, A.K. et Dubes, R.C. (1988) Algorithms for Clustering Data. Prentice-Hall. https://homepages.inf.ed.ac.uk/rbf/BOOKS/JAIN/Clustering_Jain_Dubes.pdf

[15] Esther Jebarani, P., Umadevi, N., Dang, H. and Pomplun, M. (2021) A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection. *IEEE Access*, **9**, 146153-146162. https://ieeexplore.ieee.org/document/9590488

[16] Breast Cancer Histopathological Database (BreakHis). https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/