JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

BREAST CANCER ULTRASOUND IMAGES CLASSIFICATION USING A LIGHTWEIGHT MODEL

[1]RAHAJANIAINA Andriamasinoro, [2]RATIARISON Adolphe Andriamanga

[1]Associate Professor, [2]Professor Emeritus
[1]Department of Mathematics Computer Science and Applications
[1]University of Toamasina, Toamasina, Madagascar

Abstract: As other cancer diseases, breast cancer is a serious problem of human life and especially for the women. One way to predict and know its state is the use of artificial intelligence. In this paper, we proposed a very practical solution to classify breast cancer ultrasound images that can be used on mobile devices such as smartphones. The proposed lightweight model was the combination of mobilNetV3Small and a customized version of the SSDLiteX auxiliary stage. The Balancing technical was used to perform the highest results. After more experiments, the proposed approach achieved 96% for an overall accuracy and for the global F1-score. The results saw that our model is less complex, lightweight and could be deployed on the mobile device. Moreover, the result demonstrates that our system surpassed others previous models using the same kind of the dataset.

***IndexTerms* – breast cancer disease, lightweight deep learning model, breast cancer classification.**

I. Introduction

Cancer is among the disease that kills many people every year. In 2022, 19,976,499 cases were diagnosed in the world WCRF (2024). One of the most common cancer diseases is breast cancer. It is important to note that in the majority of cases, this one affect women much more than men. According to WCRF (2024), there were 2,296,840 new cases of breast cancer in 2022 among women in the world. The breast cancer problem requires powerful equipment to perform the early diagnosis/prediction and to avoid serious cases. Or, these ones are more expensive and few hospitals have them. In fact, more specialists perform manual diagnosis; this situation could lead to the misdiagnosis and several women have had their breasts removed accidentally.

Due to the evolution of artificial intelligence, several researches are carried out in order to enhance the quality of breast cancer detection and classification tools to achieve the highest level of precision possible. Thus, Nawaz et al. (2018) designed a customized version of DenseNet model to build a breast cancer classifier using transfer learning. Their model has been contained four dense blocks and three transition layers to classify breast cancer tumors. To evaluate their model, they used the BreakHis dataset for training and testing. It has been contained 7909 microscopic biopsy images of benign and malignant breast tumors. Their results showed a multi-classification with 96% of average accuracy. In Islam et al. (2024), the authors made an evaluation of five different machine learning methods using a primary dataset (500 patients from Dhaka Medical College Hospital). They have been used five different supervised machine learning techniques such as decision tree, random forest, logistic regression, naive bayes, and XGBoost, to perform a classification of the images within the dataset into benign tumor or malignant tumor case. The final experiments found that XGBoost achieve the highest accuracy and F1-score of 96%. Moreover, ensemble deep learning model was proposed by Sami et al. (2024) for detection and classification of breast cancer images into benign and malignant class. Thus, they used the combination of GoogleNet, VGGNet, ResNet, and DenseNet as feature extractors. Then, the result has been merged using a fully connected layer to achieve classification. The ensemble deep learning model gave an F1-score of 84%. In this context, the previous researches were more complex and required more time computational. Face to this situation, we propose a lightweight model to perform classification of breast ultrasound images into two classes: malignant and benign. Our approach gave a comparable result compared to the previous work even if it is less complex and lightweight.

The rest of this study is organized as follows. A brief previous work on the different breast cancer classification approach is described in section II. Section III presents the proposed approach. Results and discussion about our approach are discussed in section IV and section V conclude this paper.

II. PREVIOUS WORK

Jasmir et al. (2018) presented classification model based on the Multilayer Perceptron algorithm. To evaluate their model, they used the datasets from the University of Medicine Center, Institute of Oncology, Ljublijana, Yugoslavia. Among of 286 data, 201 concerned No-Recurrences-Events classes, 85 corresponded of Recurrences-events classes. Some pre-processing like labelling, sorting and data cleaning technique was applied. Then, the data was divided randomly into 10 partitions to achieve the Cross Validation technique. Thus, the data fed into the model and achieved accuracy level of 96.5%.

Gami et al. (2023) worked with convolutional neural network (CNN) in order to classify cancerous cells. They used the breast cancer histopathological image classification (BreakHis) dataset collected by P&D Laboratory, Brazil. They used only 1305 samples of this dataset consisting of 407 benign samples and 898 malignant ones. Several pre-processing was applied to the dataset to ameliorate the images qualities. After training and fine tuning, their approach gave 82% of an accuracy.

A novel system for classify the breast cancer images was implemented in Sajid et al. (2023). It was a variant of VGG called compact VGG or cVGG, which was combined with HOG and LBP (called handcrafted) features to obtain the highest performance. In their study, they used Curated Breast Imaging Subset of DDSM (CBIS-DDSM) dataset Lee et al, (2017) containing 6,671 mammogram images in total. Some pre-processing was applied to the dataset to transform them to the PNG format. This one was split to 80% for training set and 20% for test one. After the evaluation of some transfer learning models such as VGG16, ResNet50, DenseNet121, InceptionResNetV1 and know their drawback, the authors performed their model using the same dataset and it achieved an accuracy of 91.5%.

Mohammed et al. (2023) proposed a deep learning-based model for classifying breast ultrasound images obtained from Baheya Hospital for Early Detection & Treatment of Women's Cancer in Cairo, Egypt. It is formed by 780 images without a mask and 1583 images with a mask. This dataset was divided into three classes as benign, malignant and normal. Their study was performed into four steps. The pre-possessing was reserved to ameliorate the images qualities and the results using color conversion format, applying thresholding to binarized the images and contours detection. Then, segmentation was applied to transform image into meaningful objects to facilitate the later process. They used U-Net architecture to achieved the segmentation task. Thus, MobileNet was used for features extraction. To achieve classification task, the data was split into 80% for train set and 20% for test set. Then, this one was fed to the variant of VGG16 model to classify into benign, malignant or normal case. After experiments, their model achieves 99.29 % of training accuracy and 89.73% of F1-Score.

Ghrabat et al. (2022) designed a fully automated model based on deep learning approach to classify tissues in a mammographic breast image. Several pre-processing was used to remove the pectoral muscle from the image to facilitate the classification task and to save computing time. 80% of 115 images were used for training and the rest for testing. They used cross-validation approach on test data including 82 images for train and the rest for validation. Face to previous work, the proposed model achieved training accuracy of 94.89% and validation accuracy of 88.32%.

Abunasser et al. (2022). Introduced the comparative study to evaluate six deep learning base models such as VGG16, Xception, ResNet50, InceptionV3, MobileNet and Breast Cancer Convolutional Neural Network (BCCNN) their model to classify breast cancer images. The BCCNN model was formed by 18 layers and it has 2.1x 107 parameters approximatively. They used the Breast "Cancer Histopathological Image Classification (BreakHis) downloaded from Kaggle website. This dataset contains 7909 images of breast cancer using various magnifying factors (40X, 100X, 200X, and 400X) divided into benign and malignant case. Among of 7909 images, 2,480 benign and 5,429 malignant. In order to overcome overfitting, they applied data augmentation using Generative Adversarial Networks (GANs) and they obtained 40,000 images. Then, a split ratio of 60:20:20 was utilized to divide the dataset into train, validation and test set. Using 4 magnifying factors, their dataset consists of 4 datasets. Each one was constituted by 10,000 images. The 4 datasets have been grouped to form one dataset. Each one was used to assess each model. Their study shows that their model performed the best accuracy of 98.30% and F1-score of 98.28%.

Ali et al. (2023) used numerous techniques to perform breast cancer classification. Thus, they used data augmentation, transfer learning and meta-learning ensemble learning to achieve their goal. The Breast Ultrasound Imaging (BUSI) dataset was employed to evaluate the performance of the proposed technique. It is divided into benign and malignant case. After data augmentation, the dataset contains 10,000 images. Some pre-processing was performed to enhance the images qualities. Moreover, the images within the dataset were split into three categories: 70% for train, 10% for validation and 20% test set. Thus, training data passed through three pre-trained model in order to extract some meta-features. Then, the meta-learner was trained using these ones on the validation and test set to make final prediction. The suggested approach achieved an accuracy of 90%.

Jafari and karami (2023) presented a method based on feature extraction and reduction to classify the breast cancer in mammography images. They used some pre-trained model as features extractors. After that, these features were reduced and fed to four machine learning such as neural network (NN), k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM) to classify into cancer or no cancer. To evaluate their approach, they used three datasets: the radiological society of north America (RSNA) dataset, the mammographic image analysis society (MIAS) and the digital database for screening mammography (DDSM). The first one contains 54,713 images in dicom format, the second one 322 mammographic images and the last one includes 55,890 images. For the RSNA dataset, the proposed model performed an accuracy of 92%. Then, for the MIAS dataset, their approach achieved 94.5% and for DDSM dataset, your model gave an accuracy of 96%.

III. PROPOSED APPROACH

we presented in this part in detail our approach. As we mentioned early, Face to the situation with previous work and in the development country, we introduced a lightweight model formed by two components: the backbone and the auxiliary stage. The backbone was based on MobilNetV3Small, which is the lighter version of MobileNetV3. Its role was the features extractor. In fact, we used the last feature extractor layer that has an output stride of 16 called C4 and the last one that has an output stride of 32 called C5 to obtain the best features. Otherwise, the auxiliary stage was inspired from the SSDLiteX auxiliary stage Kang (2023).

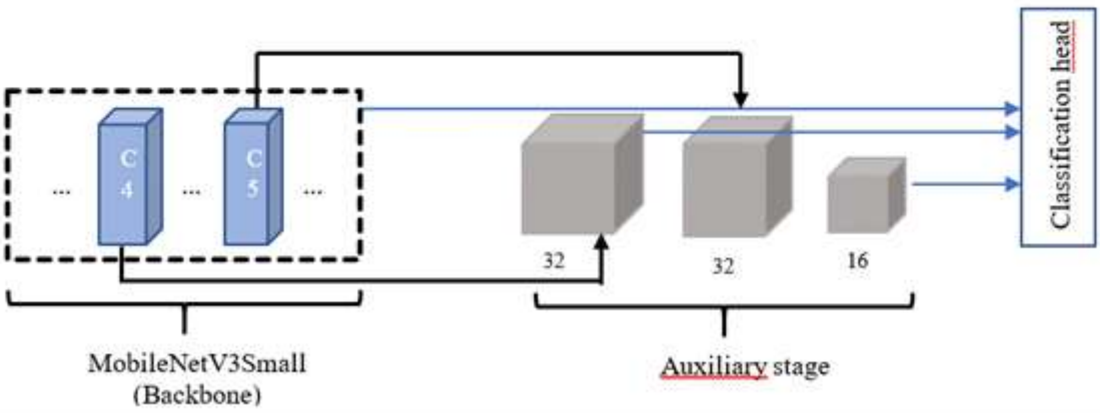

Figure 1: Proposed model architecture Andriamasinoro & Adolph (2024)

To evaluate our model, we used publicly available dataset from Kaggle web site. Our dataset contains 9015 breast ultrasound images, 70% were used for training and the rest of this one was divided equally in to validation and test set. So, the training data was balanced in order to avoid overfitting and for obtaining better results. Then, this data fed to the model to be classified into benign or malignant case. The Fig.2 shows some sample of the images in the dataset.

IV. RESULTS AND DISCUSSION

We conducted several experiments during the training step although the batch size of 1 and an epoch of 20 performed the best result. The model was implemented on Intel(R) Core (TM)i7-1255U CPU, 10 cores, 12 threads @2.30GHz, and 24 Gb RAM. We have been developed our system using python, TensorFlow, NumPy, Matplotlib, Scikit-learn, Panda, Seaborn and Keras libraries. With the intention of doing a lower variance, ameliorate the generalization, and minimize loss of the proposed model, we applied a shuffling technique with a seed of 123. We used two metrics such as the accuracy and F1-score for evaluating our approach. The accuracy gives the proportion of the correct prediction made by the model. Thus, this statistical measure performs the ration for both true positives and true negatives to the true total number cases and the F1-score correspond to the harmonic mean of the precision and recall. It gives a balance between these two metrics.

After more experiments, our model achieves 96% for an overall accuracy and the global F1-score. Compared to the previous models using the same kind dataset Mohammed et al. (2023), our model is less complex and lightweight but it provides the higher results. It shows that the proposed model is more efficacy for the breast cancer ultrasound images dataset. Fig.3 illustrates the training and validation loss, the training and validation accuracy and the training and validation F1 score.

Benign

Malignant

Figure 2: Some sample ultrasound images from the dataset

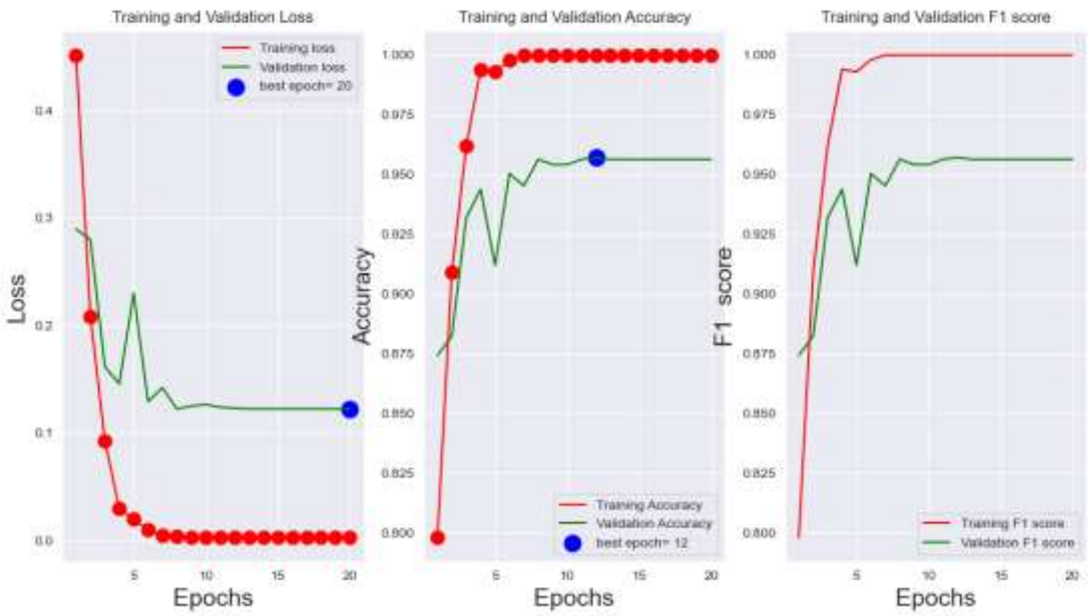

Figure 3: the training and validation loss - the training and validation accuracy - the training and validation F1 score

The confusion matrix presented in Fig.4 below illustrates the global result of the classification. It demonstrates the high performance of the proposed approach into classifying each class of breast cancer ultrasound images.

V. CONCLUSION

In this work, we have explained the result of our research about breast cancer classification. Firstly, we have studied the previous work to know the situation about this team. Thus, as we saw the majority of previous model was complex and heavier and these required an infrastructure more expensive. This situation conducts us to propose a lightweight system inspired from the SSDLiteX approach to classify breast cancer into benign or malignant using breast cancer ultrasound images. The proposed model compared with previous ones achieve a higher performance and it could run on the low resource environment and answer the need of the specialist within the hospitals.

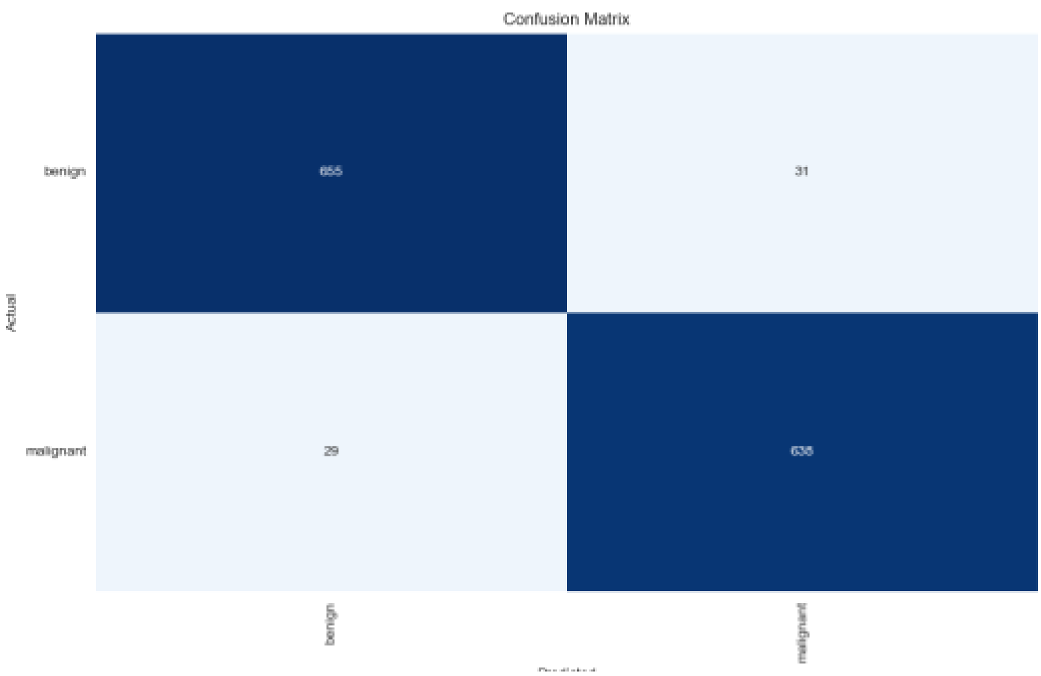

Figure 4: Confusion matrix

REFERENCES

[1] Abunasser et al. (2022). Convolution Neural Network for Breast Cancer Detection and Classification Using Deep Learning. Asian Pacific Journal of Cancer Prevention, volume 24, Issue 2: 531-544. DOI:10.31557/APJCP.2023.24.2.531

[2] Ali et al. (2023). Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks. Diagnostics 2023, volume 13. 19 pages. https://doi.org/10.3390/diagnostics13132242

[3] Andriamasinoro & Adolph (2024). Plant leaf disease classification using a lightweight model. Journal of Emerging Technologies and Innovative Research, 2024, Vol. 11, Issue 11. 225-229.

[4] Gami et al. (2023). Breast Cancer Detection Using Deep Learning. https://doi.org/10.1007/978-981-19-7982-8_8

[5] Ghrabat et al. (2022). Fully automated model on breast cancer classification using deep learning classifiers. Indonesian Journal of Electrical Engineering and Computer Science, volume 28, No. 1, October 2022: 183~191. DOI: 10.11591/ijeecs. v28.i1. 183-191

[6] Islam et al. (2024). Predictive modelling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. Scientific Reports, 2024. 17 pages. https://doi.org/10.1038/s41598-024-57740-5

[7] Jasmir et al. (2018). Breast Cancer Classification Using Deep Learning. International Conference on Electrical Engineering And Computer Science 2018: 237-242.

[8] Jafari and karami (2023). Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. Information, volume 14. 14 pages. https://doi.org/10.3390/info14070410

[9] Kang (2023). SSDLiteX: Enhancing SSDLite for Small Object Detection. Applied sciences, 2023, 13. 9 pages. https://doi.org/10.3390/app132112001

[10] Lee et al, (2017). A curated mammography data set for use in computer-aided detection and diagnosis research, Scientic Data (2017).

[11] Mohammed el al. (2023). Breast Cancer Detection Using Deep Learning Technique Based On Ultrasound Image. https://arxiv.org/pdf/2312.05261

[12] Nawaz et al. (2018). Multi-Class Breast Cancer Classification using Deep Learning Convolutional Neural Network. International Journal of Advanced Computer Science and Applications, Volume 9, N° 6, 2018: 316-322.

[13] Sajid et al. (2023). Breast Cancer Classification using Deep Learned Features Boosted with Handcrafted Features. arXiv:2206.12815v2 [eess.IV].

[14] WCRF (2024). Breast Cancer Statistic. https://www.wcrf.org/preventing-cancer/cancer-statistics/breast-cancer-statistics/ Accessed on 21/12/2024 at 21:18 p.m.