International Journal of Internet, Broadcasting and Communication Vol.16 No.1 124-131 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.1.124

IJIBC 24-1-15

Breast Cancer Classification in Ultrasound Images using Semi-supervised method based on Pseudo-labeling

Seokmin Han

Professor, Major of Data Science, Korea National University of Transportation, South Korea
E-mail: seokmin.han@ut.ac.kr

Abstract

Breast cancer classification using ultrasound, while widely employed, faces challenges due to its relatively low predictive value arising from significant overlap in characteristics between benign and malignant lesions, as well as operator-dependency. To alleviate these challenges and reduce dependency on radiologist interpretation, the implementation of automatic breast cancer classification in ultrasound image can be helpful. To deal with this problem, we propose a semi-supervised deep learning framework for breast cancer classification. In the proposed method, we could achieve reasonable performance utilizing less than 50% of the training data for supervised learning in comparison to when we utilized a 100% labeled dataset for training. Though it requires more modification, this methodology may be able to alleviate the time-consuming annotation burden on radiologists by reducing the number of annotation, contributing to a more efficient and effective breast cancer detection process in ultrasound images.

Keywords: *deep learning, semi-supervised method, breast cancer, classification.*

1. INTRODUCTION

Breast cancer detection using ultrasound, while widely employed, faces challenges due to its relatively low predictive value arising from significant overlap in characteristics between benign and malignant lesions, as well as operator-dependency [1]. To Alleviate these challenges and reduce dependency on radiologist interpretation, the implementation of automatic breast cancer classification in ultrasound images can be helpful. The implementation of automatic breast cancer classification serves as an objective reference, enhancing the real-time nature of breast ultrasound imaging and providing timely information to support radiologists in making more informed decisions. The integration of automated processes not only reduces the time spent on the visual inspections but also enhances overall working efficiency. Recently, deep learning methods have gained attention for their potential in early breast cancer detection [2-4]. However, developing an automatic breast cancer detection method based on deep learning method requires annotations for each training image, which is a time-consuming task for radiologists [5]. To deal with this problem and train neural networks on

Manuscript Received: December. 22, 2023 / Revised: January. 13, 2024 / Accepted: January. 23, 2024
Corresponding Author: seokmin.han@ut.ac.kr
Tel: +81-31-460-0587
Professor, Major of Data Science, Korea National University of Transportation, South Korea

incompletely annotated data, we propose a semi-supervised deep learning framework for breast cancer detection. In this approach, the neural network learns to classify unlabeled data with a limited set of annotated training data and subsequently assigns pseudo-labels to the unlabeled data [6-8]. This method aims to alleviate the time-consuming annotation burden on radiologists, contributing to a more efficient and effective breast cancer detection process in ultrasound images [9].

2. EXPERIMENTS

The data used in this research is Breast Ultrasound Images(BUSI) dataset, and it includes breast ultrasound images among women in ages between 25 and 75 years old [10]. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500 × 500 pixels, and resized to 255× 255 in this research. The images are categorized into three classes, which are normal, benign, and malignant. The number of images in each class is 487 benign images, 210 malignant images, 133 normal images. Figure 1 illustrates the overall concept of this study. Initially, around 30% of the total data, 240 images, are randomly selected as the test set. This test set does not participate in the training process. For the remaining 70%, 540 images, annotations are provided for some of the data to create labeled data, leaving the rest as unlabeled training data. The labeled data is first used for training a ResNet50 as the backbone neural network [11].

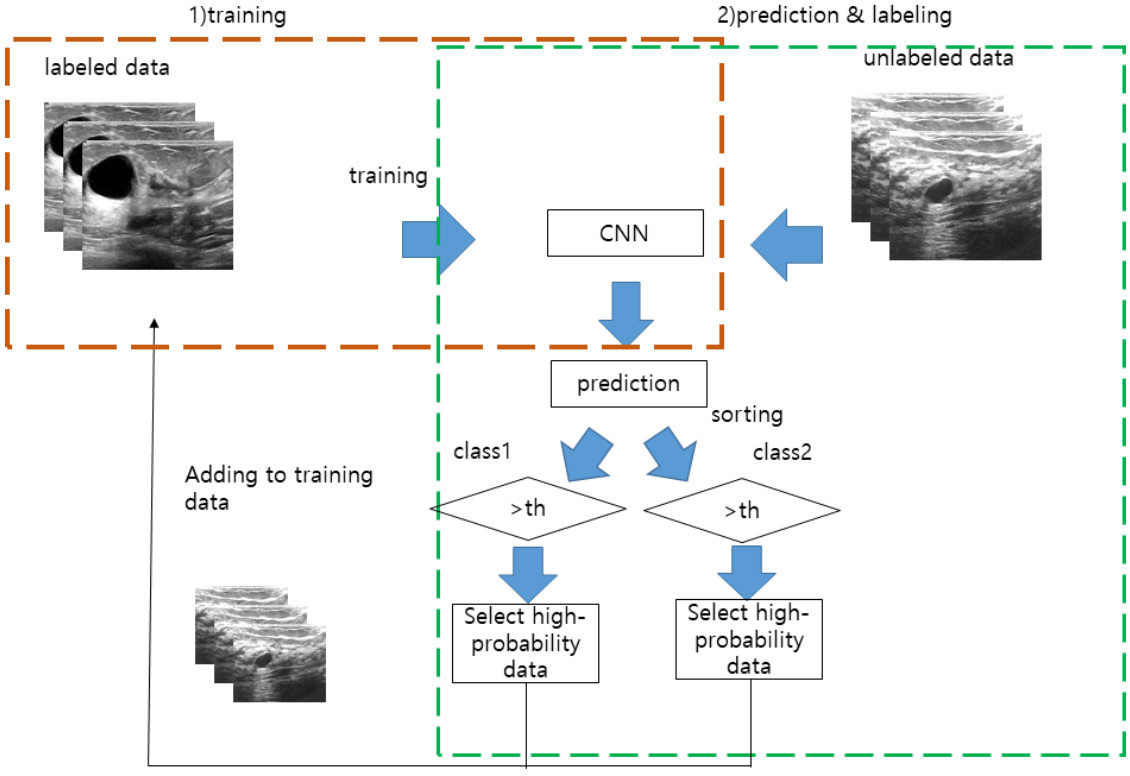

Figure 1 Overview of the proposed method

The neural network trained with the labeled images is then fed with 1/5 of the unlabeled data to obtain prediction results. The reason for selecting 1/5 is to perform pseudo-labeling on the entire unlabeled data in 5 steps. The obtained prediction results represent the probabilities corresponding to each class. For example, for

the benign class, the prediction results are sorted in descending order, and values greater than the threshold of 0.6 can be considered as benign class. It is assumed in this research that the ratio of each class is known. Benign class has ratio of 50%, malignant class 35%, and normal class 15%. Therefore, for the benign class, the top 50% of the sorted data, whose prediction values are higher than the threshold, are considered as benign class. The same process is also operated for non-benign class. In this research, we consider 3 two-class problems(benign vs non-benign, malignant vs non-malignant, normal vs non-normal). For malignant class, the top 35% of the sorted data, larger than the threshold, are considered as malignant. For normal class, the top 15% of the sorted data as normal. Pseudo-labeling is then performed according to the selected data. The pseudo-labeled data, combined with the existing labeled data, forms a new training dataset, and the neural network is retrained. This process is repeated over 5 steps, allowing the neural network to complete its training on all data. The pseudo-code for this process is summarized as follows:

Select sample training data randomly.
Annotate the selected sample data.
Separate the unlabeled data into 5 groups.
Repeat for 5 loops:

Train the neural network with the selected data with annotation.
Predict on one unlabeled data group.
Sort the result.
Select those predicted result larger than the threshold
Add the predicted data to the training data.

---

3. RESULTS

Figure 2 provides examples of BUSI images used in this study. The left image is benign, the middle image is malignant, and the right image is normal.

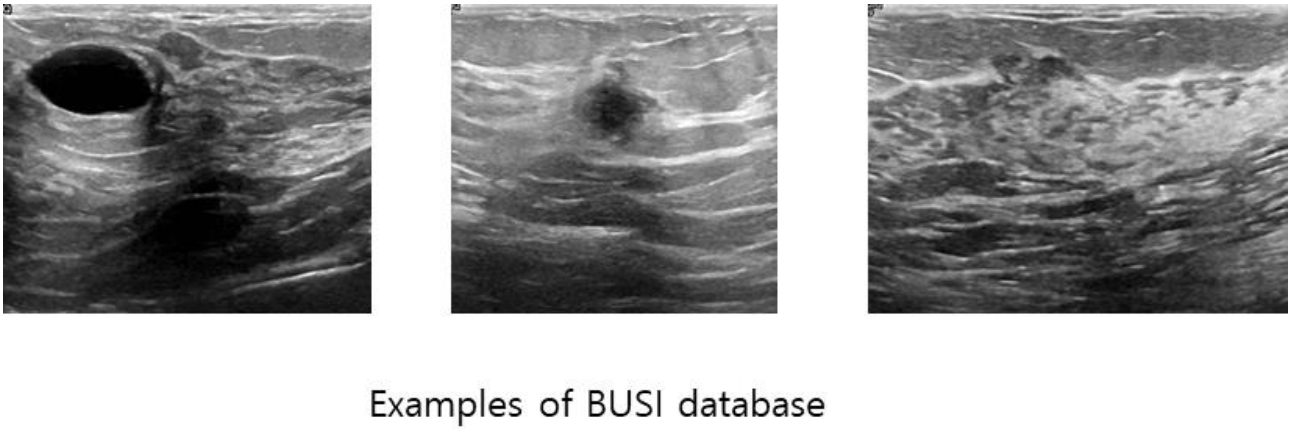

Figure 2. Examples of BUSI database. (left)benign,(middle)malignant,(right)normal

The system used for experiments is based on an Intel Core i7-13700 PC with an attached RTX4090. The ResNet50 is employed as the backbone neural network. Cross-entropy is used as the loss function, with a

learning rate set to 0.0001, and the Adam optimizer applied. The model's structure is depicted in Figure 3. As can be seen in Figure 3, Global average pooling is attached to the backbone network, and fully connected(FC) layer follows for classification. In this research, ResNet50 is used as the backbone network, pre-trained using ImageNet.

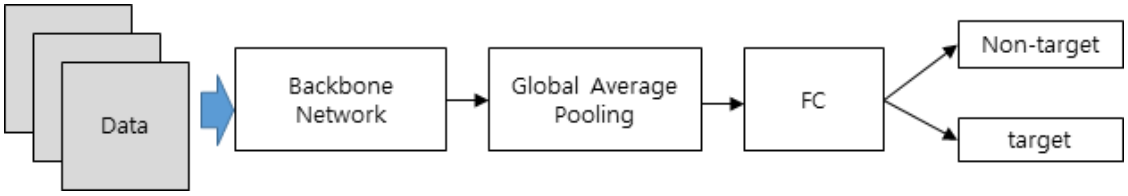

Figure 3. the structure of the used neural network model

Different numbers of initial data are used to train the backbone Neural Network after randomly selected within the training data. This is done to observe the performance change of the model over the number of the initially labeled data. The remaining data is denoted as unlabeled. The neural net trained with the initially labeled data was used to predict on the unlabeled data. As mentioned in the Experiment Section, the unlabeled data is sorted in descending order for each class, whose values should be larger than the threshold. Considering the ratio of each class, some of the sorted data were selected. Those selected data were added to the corresponding class in the training data set. If a predicted value is not larger than the threshold, it was not added. In this way, pseudo-labeling was performed based on the prediction results. After pseudo-labeling, the Network model was trained on the initially labeled data and the added data. This process is repeated in 5 cycles, for the unlabeled data were divided into 5 groups. After each cycle, predictions are made on the test data to investigate the performance change over the number of the pseudo-labeled data. Training is conducted without using a validation set, ending just after 20 epochs, at which point predictions were made on the test data. Under the same conditions, this experiment was conducted five times, and averages and standard deviations are calculated.

The results of the experiments varying the number of the initial labeled data for benign vs non-benign are shown in Table 1 and graphically represented in Figure 4. As can be seen in Table 1 and Figure 4, training process begins with 51 labeled data, and it achieves about 82% accuracy.

Table 1. Benign vs non-Benign case

Labeled data	51	149	251	349	451
Average	0.6942	0.7625	0.8050	0.8217	0.8217
Std	0.0563	0.0281	0.0418	0.0344	0.0380
Max	0.7542	0.8042	0.8500	0.8625	0.8583
Min	0.6083	0.725	0.7583	0.7875	0.7667

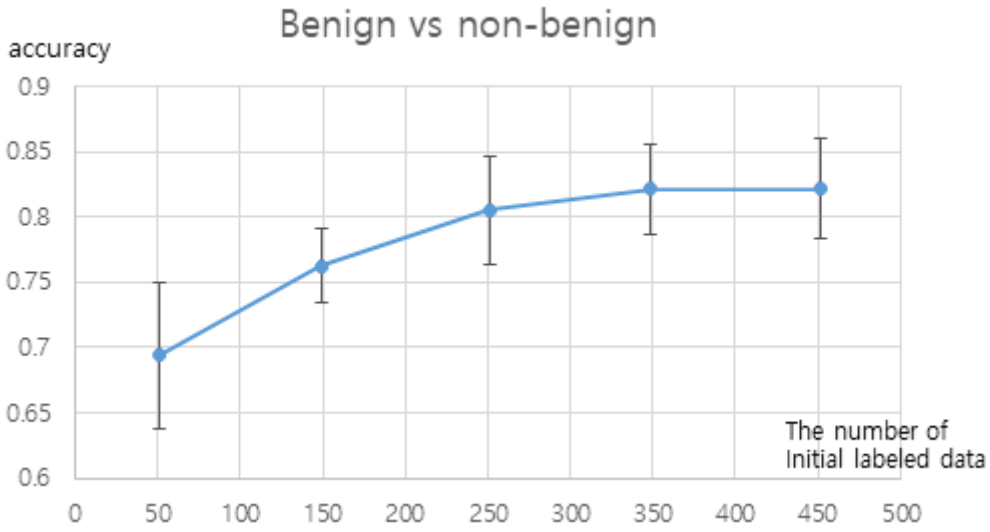

Figure 4. the effect of the number of the initial labeled data for benign vs non-benign case

The results for malignant vs non-malignant are shown in Table 2 and in Figure 5. As can be seen in Table 2 and Figure 5, it achieves about 86% accuracy.

Table 2. malignant vs non-malignant case

Labeled data	51	149	251	349	451
Average	0.8042	0.8778	0.8847	0.8611	0.8681
Std	0.0356	0.0241	0.0087	0.0856	0.0214
Max	0.8375	0.8917	0.8917	0.9167	0.8917
Min	0.7667	0.8500	0.8750	0.7625	0.85

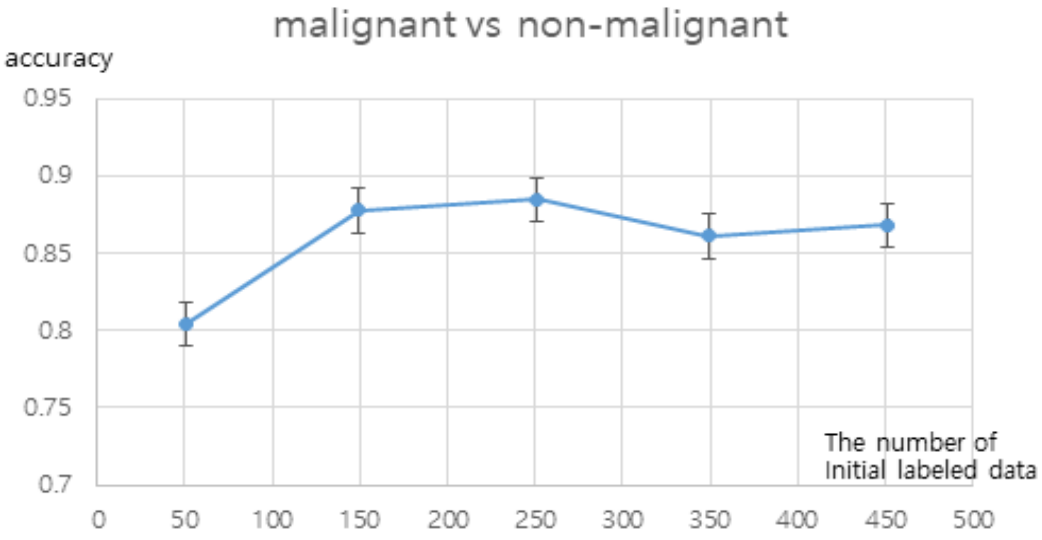

Figure 5. the effect of the number of the initial labeled data for malignant vs non-malignant case

The results for normal vs non-normal are shown in Table 3 and Figure 6. As can be seen in Table 3 and Figure 6, it achieves about 92% accuracy.

Table 3. normal vs non-normal case

Labeled data	51	149	251	349	451
Average	0.8097	0.8847	0.9097	0.9153	0.92913
Std	0.0741	0.0349	0.0315	0.0121	0.0144
Max	0.8750	0.9250	0.9458	0.9292	0.9458
Min	0.7292	0.8625	0.8875	0.9083	0.9208

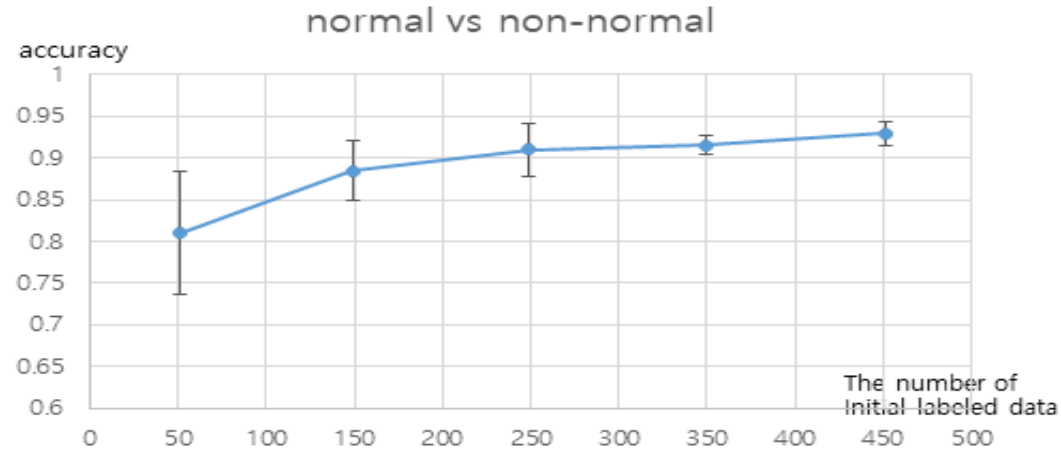

Figure 6. the effect of the number of the initial labeled data for normal vs non-normal case

4. DISCUSSION

In the proposed method, the experiments demonstrate that training with 251 labeled data could be the very balanced point between the final performance and the amount of labeling. Theoretically, when all the training data were labeled, it can achieve the best performance. However, we also have to consider the time and the cost for 100% labeled training dataset, which is fully-supervised. Therefore, in this research, providing the labeling with about the half of the total training data(540 images) could achieve a reasonable level of performance. The initial data size has a significant impact on the final performance, especially when the number of initial labels is low. As a refence, the model achieved a performance of an average accuracy of 85.8%, 89.0%, and 93.9% for benign, malignant, and normal case, respectively, when trained on all 540 labeled data. The primary focus of this research was to examine the impact of initial data size, ultimately achieving an accuracy of close to 90%. Though it still requires modification, it achieved acceptable performance. Combining with methods like Contrastive Learning could lead to additional performance improvements. Moreover, future efforts should be directed towards applying the proposed model to data beyond BUSI data [12, 13].

5. CONCLUSION

In this study, a model trained through semi-supervised learning was utilized to classify BUSI data base. During the semi-supervised learning process, the model performed pseudo-labeling based on the predictions of a certain amount of data. Following pseudo-labeling, a thresholding and sorting process were operated, and these pseudo-labels were added back to the training data, gradually improving the prediction performance. We proposed a semi-supervised deep learning framework for breast cancer classification. In the proposed method, we could achieve reasonable performance utilizing less than 50% of the training data for supervised learning in comparison to when we utilized a 100% labeled dataset for training. Though it requires more modification, this methodology may be able to alleviate the time-consuming annotation burden on radiologists by reducing the number of annotations, contributing to a more efficient and effective breast cancer detection process in ultrasound images.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1049467) and by Korea National University of Transportation in 2023.

References

[1] Y. Ouyang, et al., “Classification of Benign and Malignant Breast Tumors Using H-Scan Ultrasound Imaging,” Diagnostics (Basel), Vol. 8, 2019. DOI:https://doi:10.3390/diagnostics9040182.

[2] S. Han, et al., “A Weak and Semi-supervised Segmentation Method for Prostate Cancer in TRUS Images,” Journal of Digital Imaging, Vol. 33, 2020. DOI:https://doi.org/10.1007/s10278-020-00323-3

[3] L. Shen, et al., “Deep Learning to Improve Breast Cancer Detection on Screening Mammography,” Scientific Reports, Vol. 9, 2019. DOI:https://doi.org/10.1038/s41598-019-48995-4

[4] S. Han, et al, “A Practical Implementation of Deep Learning Method for Supporting the Classification of Breast Lesions in Ultrasound Images,” International Journal of Advanced Smart Convergence Vol.8 , pp. 24-34, 2019. DOI:https://doi.org/10.7236/IJASC.2019.8.1.24

[5] Y. Gu, et al., “Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study,” Insights into Imaging, Vol. 13, 2022. DOI:https://doi.org/10.1186/s13244-022-01259-8

[6] X. Yang, et al, “A Survey on Deep Semi-Supervised Learning,” IEEE Transactions on Knowledge and Data Engineering, Vol. 35, 2022. DOI:https://doi.org/10.1109/TKDE.2022.3220219

[7] E. Arazo, et al, "Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning," proceedings of International Joint Conference on Neural Networks (IJCNN), 2020. DOI:https://doi.org/10.1109/IJCNN48605.2020.9207304.

[8] Y. Chen, et al, "Semi-Supervised Learning under Class Distribution Mismatch," Proceedings of the AAAI Conference on Artificial Intelligence, 2020. DOI:https://doi.org/10.1609/aaai.v34i04.5763

[9] S. Lee, S. Han, "Detection Fastener Defect using Semi Supervised Learning and Transfer Learning," Journal of Internet Computing and Services, Vol. 24, 2023. DOI:https://doi.org/ 10.7472/jksii.2023.24.6.91

[10] W. Al-Dhabyan,et al, "Dataset of breast ultrasound images," Data in Brief, Vol. 28, 2020. DOI:https://doi.org/10.1016/j.dib.2019.104863

[11] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016. DOI:https://doi.org/10.1109/CVPR.2016.90

[12] S. Lee, and S, Han, "Detection Fastener Defect using Semi Supervised Learning and Transfer Learning," J. Internet Comput. Serv., Vol. 24, pp. 91-98, 2023. DOI: http://dx.doi.org/10.14801/jkiit.2022.20.2.

[13] T. Chen, K. Simon, N. Mohammad and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," In International conference on machine learning, pp.1597-1607, 2020. DOI: https://arxiv.org/abs/2002.05709.

[14] X. Wang, G. Qi, "Contrastive Learning With Stronger Augmentations," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, pp..21271-21284, 2022. DOI:https://doi.org/10.1109/TPAMI.2022.3203630