Self-supervised Visual Transformers for Breast Cancer Diagnosis

Nurbek Saidnassim*†, Beibit Abdikenov*†, Rauan Kelesbekov*, Muhammad Tahir Akhtar*, Prashant Jamwal*

*School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan, †Relive Research LLP, Astana, Kazakhstan

E-mails: {nurbek.saidnassim, beibit.abdikenov, rauan.kelesbekov, muhammad.akhtar, prashant.jamwal}@nu.edu.kz

***Abstract*— Owing to the growing incidences, breast cancer is considered the world's most prevalent cancer both in terms of morbidity and mortality rates. While breast cancer treatment is quite effective, higher mortality is caused due to the delayed diagnosis. Mammogram image processing using artificial intelligence techniques can be promising in the early diagnosis and Convolutional Neural Networks (CNN) have shown some potential in image classification and semantic segmentation. However, working with unlabeled image data poses challenges and while the manual labeling is not efficient, pre-trained CNNs also perform poorly on the actual medical images. In this paper, we propose to use transformer-based networks in the field of computer vision. Previously published research also suggests a predominance of transformer-based vision models over the convolutional models. Bootstrap Your Own Latent (BYOL), which is Self-supervised learning (SSL) algorithm, is proposed in this paper to process mammograms images for diagnostic purposes. BYOL can handle unlabeled image data and improve the deep learning mechanism for data analytics. UNet networks with Xavier initializations have been considered for the baseline to compare with their performances when implemented with and without BYOL. It is shown that the overall network performance improves when they are trained using the proposed BYOL algorithm.**

Keywords— Mammogram image processing, UNet-CNN, style, Bootstrap Your Own Latent (BYOL)

I. Introduction

Cancer is one of the predominant triggers of female deaths worldwide and claims more lives than malaria, tuberculosis, or cardiovascular diseases. According to the International Agency for Research on Cancer (IARC) and the American Cancer Society, new cancer cases, across the world, in 2018 were as many as 17.1 million [1]. As per the recent assessments, the cancer incidences may reach up to 27.5 million by 2040, besides an estimated 16.3 million fatalities as a consequence [1]. Among all the cancer types, breast cancer is more prevalent in women, and according to the IARC report, breast cancer incidences account for 25% of the cancer occurrences in women worldwide. Developing countries, housing 82% of the world population, record 53% of the breast cancer incidences [1]. Breast cancer is also the leading cause of mortality among women and is only second to lung cancer [1]. Some of the main triggers for breast cancer are hormonal, lifestyle, and environmental changes [2, 3]. The cancer cells further spread from breasts to lymph nodes rather quickly and so an early diagnosis is also one of the remedies and is important. There are indicators for breast cancer such as malfunctioning of milk ducts, growth of glandular tissues, etc. However, a low percentage (20%-50%) of patients in the low and middle income countries (LMIC) are identified in the earlier stages [4]. The present protocol suggests breast cancer screening with radiographic and clinical evaluations. A three-dimensional mammography called breast tomosynthesis, is an advanced type of breast imaging which is basically a computer reconstruction of x-rays images of the breast. Breast tomosynthesis aids in the early detection and diagnosis of breast cancer even before the symptoms are experienced [5-8]. Acquiring mammograms require low doses of radiation and is a safe procedure compared to the previous approaches such as Ultrasound and magnetic resonance imaging (MRI) [9].

In the field of healthcare computer vision, the most pressing issue is the scarcity of data. It is quite difficult to find sufficient number of images to train an adequately accurate model. In classical computer vision, one can use a pretrained network, in order to simplify the search of optimal weights, giving the network the initial head start. Previously, two way classification of breast cancer images has been used for the prognosis purpose [10]. Review on the computer-aided breast cancer detection can be accessed from [11-14]. Automated Detection of breast tumor in different imaging modalities has been discussed for the improved performance of computer added diagnosis (CAD) systems for diagnosis/detection of breast masses [15]. A review of the computer-aided detection and diagnosis of breast cancer in digital mammography can be accessed from [16]. However, most of the images are not apposite for the application of medical image analysis. Thus, using ImageNet pretrained weights may not be the best option. As a solution to this problem, self-supervised learning, which is essentially a representation learning, has been proposed in the literature [17, 18]. During self-supervised learning, a model tries to learn a representation of data without any labels. Instead, it uses certain data transformations as labels and fits a model on them. Self-supervision has been successfully applied to power such algorithms as Bidirectional Encoder Representations from Transformers (BERT) for natural language processing [19], cancer prognosis from mammography images [20]. The rest of this paper is organized as follows: Related previous work has been discussed in Section II. Bootstrap Your Own Latent (BYOL) algorithm for breast cancer diagnosis has been explained in Section III with respect to its implementation on the mammogram image processing. The conduct of experiments has been elaborated in Section IV and the results from these experiments are presented and discussed in Section V.

II. Related work

A. Breast Image Data and Preprocessing

In order to detect early signs of breast cancer, Mammograms are obtained and analyzed radiographically. Two views of each of the breasts, that are sufficient, are called, craniocaudal and mediolateral oblique (MLO) view. To avoid false-positive (FP) and false-negative (FN) cases, it is preferred that the mammograms be preprocessed before using as the training data. The preprocessed mammograms are later used for the necessary CAD system modelling [21]. Basically, the objects that are undesired such as annotations, background noises, and the manually put labels are removed during the preprocessing stage of the mammograms. Preprocessing helps in increasing the efficiencies of the predictive or the prognosis models by localizing the search region. There is another challenge in the detection of the lesions which is encountered due to the presence of the pectoral muscle (PM). These muscles mostly interfere with the MLO views and therefore it is necessary to define the contours of the PM in the breast region [21]. Removal of the PM contours from the mammograms will enhance the diagnostic precision, avoid false detection, and will eventually save the computational time required to process images [22].

B. Enhancement of Mammograms

In order to improve on the quality of mammograms image enhancement techniques are employed by which the readability can be enhanced by tuning the amount of color or grayscale differentiation existing between various image features. The techniques employed can help to detect the abrasions or lesions in the mammographs that has poor visibility and also help to improve them by tuning their contrast. Especially the mammograph features with low contrast are needed to be improved. Such regions of low-contrast containing small-scale malformations are often obscured within surrounding soft tissues and eventually lead to an incorrect diagnosis. The CAD system and the human observer can read small-scale malformations in the mammographs with ease as a consequence of these image processing performed. Furthermore, caution is needed while using these techniques as it may result in to disfiguring or deforming the anatomical structures. However, with the advent of digital imaging equipped with adjustment settings for contrast and dynamic range (the tonal difference between the lightest and the darkest parts of an image) the mammograph adjustment can be obtained with ease. There are three classes of image enhancement techniques used for the mammograms such as, techniques pertaining to frequency domain, spatial domain, together with techniques for a combination of frequency and spatial domain techniques [23, 24]. While the mammograms are digitized these days, most of the image enhancement procedures are conducted automatically during such digitization.

The above mentioned enhancement techniques can be implemented using the conventional methods, however, these methods may also increase the noise factor and therefore not preferred for the CAD based analyses [25]. There are certain region-based methods which are normally used for the contrast enhancements and for increasing visibility of the anatomical features of certain regions of interest (ROIs). These methods can be used to augment the details pertaining to the anatomy of the ROIs without the use of any artifacts. Dense breast tissue, appearing as microcalcification, are sometimes difficult to observe on a mammogram through imaging and this makes the diagnosis of breast cancer all the more difficult. The region-based methods can appropriately resolve this issue as well.

On the other hand, mammograms with microcalcification can be processed using the feature-linking models for image enhancement. Functions such as wavelets can be borrowed from the theory of signal processing that may help in recovering weak signals removing the noise. Such multiscale transforms are effective owing to their dilation and translation properties that can be suitable for nonstationary signals such as mammograph images. Applying wavelet threshold denoising method, it is possible to remove low frequency noise and preserve the frequencies above a threshold. Recently a fuzzy based enhancement technique was proposed for the contrast enhancement and noise suppression of the normalized mammograms and it was shown to be effective for the enhance the mass contours and feature identification [26].

C. Segmentation of the Mammographic Mass

Segmentation of the mammographs is essential and it needs to be well focused and precise for the detection of abnormal tissues in the breast through feature extraction. Segmentation is also important in extracting an ROI for the identification and distinction between breast regions with abnormalities and regularities. During the segmentation of mammographs, the breast region is separated from the background and the surrounding other objects. Furthermore, the mammograph may required to be partitioned into several distinct regions with target mass lesions.

We expect high false positive (FP) detection at this stage owing to higher sensitivity rate. Different approaches towards the breast mass classification have been suggested in the literature that include automatic and/or ensemble-based segmentation and classification algorithms [27]. Ensemble of several techniques helps in reducing the false positive cases at the detection stage. As mentioned above, the techniques used for the segmentation may use thresholding criterion, region-criterion, or feature and edge criterion [28].

D. Conventional Techniques for Feature Extraction

While employing machine learning based models the most informative and substantial features are learned which are later used as discriminators during the segmentation or the classification stage. Clinicians based on their prior experience and knowledge are expected to manually label these informative features which are the target domain. Manually labelling the features remains the weakest link in the chain and poses as a bottleneck. Commonly, experts use the principal component analysis (PCA), filtering techniques such as chi-square test, linear discriminant analysis (LDA), and similar other feature reduction methods. Here it is important to avoid overfitting and to select the features that are most discriminatory in nature so that any kind of redundancy in the feature space can be avoided [29]. Based on the feature characteristics, the workspace is normally divided into three regions of interest namely, morphological features identifying the shapes or the geometry of segments, texture or statistical features, and multi-resolution features.

E. Classification of Mammographic Images

Initial classification of the mammographic images is carried out in order to establish whether the lesion being analyzed falls under a normal or a cancerous region. A further classification may be required to be carried out in the event of finding a cancerous region. An advanced classification normally is conducted to determine the pathology of the cancer that helps in deciding between the malignant or the benign nature of the tissues. The accuracy of classification

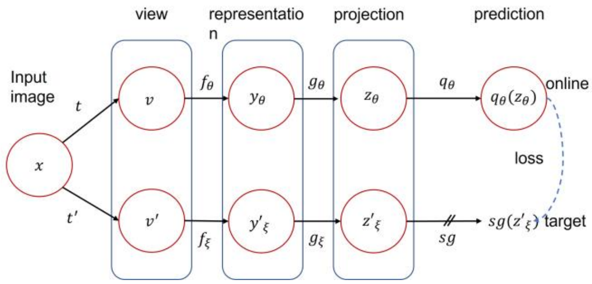

Figure 1: Bootstrap Your Own Latent

greatly depends on the previous steps of segmentation and feature extraction. Few of the important classifiers that are implemented for the breast cancer classification, include artificial neural network (ANN), support vector machine (SVM), binary decision tree, k-nearest neighbor (KNN), and simple logistic classifier [30].

F. Mammogram Classification using Convolutional Neural Networks

The convolutional neural network (CNN) is a type of deep-learning approach that is specifically applied to the image data classification. Recently, CNN-based algorithms have become quite popular owing to their capabilities in the classification of large-set images construct used in image classification. Different classifiers that are based on CNNs have been proposed in the literature. These classifiers are normally evaluated on the basis of their performance relative to the truth values generated by the histology results from biopsy and mammogram analysis by expert radiologists [31]. The CNN models, after being constructed, need to be optimized through data augmentation for better accuracies and transfer learning capabilities. It has been established that CNNs have a great potential for automatic breast cancer detection using mammograms [32]. However, working with unlabeled image data in CNNs poses challenges and while the manual labeling is not efficient, pre-trained CNNs also perform poorly on the actual medical images. Instead, we propose to use transformer-based networks in the field of mammogram analysis. Previously published research also suggests a predominance of transformer-based vision models over the convolutional models [33]. Bootstrap Your Own Latent (BYOL), which is Self-supervised learning (SSL) algorithm, is proposed in this paper to process mammograms images for diagnostic purposes. BYOL, that can handle unlabeled image data and improve the deep learning mechanism for data analytics, has been explained in the following section with its implementation [34].

III. BYOL FOR BREAST CANCER DIAGNOSIS

Bootstrap Your Own Latent (BYOL) is an approach to handle unlabeled image data and improve the deep learning mechanism for computer vision [35]. The raw image data is required to be labelled and manually assigning labels is a time consuming activity. Nevertheless, self-supervised learning, which is a nascent sub-field of deep learning, can still be used to learn from unlabeled samples. Self-supervised learning (SSL) can be implemented using either a *generative* or a *contrastive* learning approach. While *generative* learning, that uses GANs is computationally expensive, the *contrastive* approach is much less expensive and simple. Further, the

Figure 2: Mammogram and the mask after preprocessing

contrastive methods are sensitive to the choice of image augmentations and may result in *data bias* which is yet another big issue in machine learning. The proposed BYOL does not depend on negative sampling (dissimilar representations) and therefore will not have data bias and is also computationally less expensive [34]. The BYOL is a simple approach whereby the goal is to train a model so that similar samples may have similar representations. In the present work, we used BYOL self-supervision method [36] for segmentation. As has been explained, BYOL is a non-contrastive representation framework that does not rely on negative pairs to learn the representation of the dataset. It consists of two networks, namely, the online network and the target network. The target network is used to provide targets to train the online network. The BYOL method is pictorially explained in Figure 1, where the two transformations t and t' are used to transform an input $'x'$. The transformations were implemented using *Kornia*, which is a differentiable computer vision library for PyTorch (a library for deep learning on images). The model f with weights θ is used as the online network to output the representation y. The target network, similarly, has the same architecture as the online network but a different set of weights ξ is used there to provide targets to train the online network. To begin with, BYOL provides two augmented views, $v \triangleq t(x)$ and $v' \triangleq t'(x)$, from given image x. These views are obtained by applying image augmentations $t \sim \tau$ and $t' \sim \tau'$, where τ and τ' are the two distributions used for the image augmentations. Later, the online network produces a representation y_θ, together with a projection z_θ, and a prediction $q_\theta(z_\theta)$from the first view $v \triangleq t(x)$. Likewise, the target network produces y'_ξ as output and a target projection z'_ξ from the second view $v' \triangleq t'(x)$. Eventually, the Loss which is the mean squared error between the $L2$-normalized prediction $\overline{q_\theta}(z_\theta)$ and the target projection $\bar{z}'_\xi$ is calculated using (1). Intuitively, this method minimizes the distance between different variations of the similar images [35].

$$L_{\theta,\xi} \triangleq \left\|\overline{q_\theta}(z_\theta) - \bar{z}'_\xi\right\|_2^2 = 2 - 2.\frac{\langle q_\theta(z_\theta), z'_\xi\rangle}{\|q_\theta(z_\theta)\|_2 \left\|z'_\xi\right\|_2} \quad (1)$$

Here, operator $\langle .,.\rangle$ is used for the inner product, and another loss function $L'_{\theta,\xi}$ is devised to normalize the loss ($L_{\theta,\xi}$) by inducing v' to the online network and v to the target network. Therefore, the final loss function L^{BYOL} becomes the sum of the two loss functions $L'_{\theta,\xi}$ and $L_{\theta,\xi}$. During the run of the BYOL algorithm, at each training epoch, the total loss function (L^{BYOL}) is minimized varying the online network weights θ and on the other hand, the target network weights ξ are derived as the moving exponential average of $\theta: \xi \leftarrow \tau\xi + (1-\tau)\theta$, taking τ as the target decay rate.

IV. Experiments, Rsults and Discussion

A. Preprocessing

Mammography image data of patients, in the format of DICOM files, was obtained from the University Medical Center (UMC) Hospital, Nursultan, Kazakhstan. Data cleaning was performed since the mammograms had some unnecessary artifacts and labels. Later, these images were converted to the PNG bitmap image format. Since the network should receive inputs of the same size, the image data was converted to a square size before inputting to the transformations.

The image data was also cleaned by removing the artifacts such as writings on a mammogram etc. and enhancing the contrast of the images. Since, for every abnormality in the image, there has to be a separate mask, we had to scale masks size to match the input and combine each mask to the corresponding input.

B. Training

As a baseline, we first performed a supervised training of the mammographic images, and then measured and compared the accuracy with our BYOL trained model. We have selected U-Net for the base model which is a convolutional neural network and was developed for biomedical image segmentation [37]. There are many variants of UNet available on the Keras platform (open source Python library) and the one we have chosen for the present work is called UNet (ResNet34) [38]. We have also compared the results of the Mammogram image data training obtained from UNet (ResNet34) with SSL. All the models were trained using Adam optimizer (stochastic gradient descent algorithm for training models) with the 0.001 learning rate and a 0.000005 weight decay. The learning rate is reduced by the factor of 10 if the loss does not improve for 10 consecutive epochs. A weighted linear combination of Binary Cross Entropy, Dice and Focal losses with the weights of 3.0, 1.0 and 2.0 respectively were used. All models were trained for 50 epochs. For Self-Supervision, we used BYOL method with the addition of random rotations within the boundary of 10 degrees. We pretrained the ResNet model for 500 epochs on the unlabeled dataset of 1227 images.

C. Dataset and evaluation

Mammography image data of patients, in the format of DICOM files, was obtained from the University Medical Center (UMC) Hospital, Nursultan, Kazakhstan. The dataset was further divided into three parts to be used for training, validating, and testing the networks.

While 1000 mammograms were used to train the models, 231 and 340 images were used for the validation and testing, respectively. Performance of UNet (ResNet34) was used as a benchmark to compare and show improvements in the performance while using the proposed BYOL based models.

D. Results and Discussions

As a result the mean Intersection over Union (IOU) performances of the UNet (ResNet34) was enhanced, using SSL, by close to 4% and the Loss function was reduced by 12.5%. These results apparently indicate the efficacy of the Self-supervised learning (SSL) over pretrained models. During the validation stage of experiments, performance of the UNet model is increased by 16.66% after switching to BYOL.

V. Conclusions

A transformer-based networks has been developed during this research for Mammogram image processing and analysis. Bootstrap Your Own Latent (BYOL), which is a Self-supervised learning (SSL) algorithm, was designed to process mammograms images for diagnostic purpose. The results from the experiment show that the self-supervision provides better loss convergence and the improved segmentation. Future work in this direction shall be conducted to evaluate the transferability of SSL based models on different datasets of mammogram images. Using another self-supervised technique that is more focused on rotational invariance shall also be investigated.

Acknowledgment

This work was supported by the Faculty Development Competitive Research Grants Program of Nazarbayev University under the grant number 110119FD4525 (SOE2019007) and under the grant number 021220FD0251 (SEDS2021002).

References

[1] A. C. Society. (2018) Global Cancer: Facts & Figures. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition.pdf

[2] T. Blakely, C. Shaw, J. Atkinson, R. Cunningham, and D. Sarfati, "Social inequalities or inequities in cancer incidence? Repeated census-cancer cohort studies, New Zealand 1981-1986 to 2001-2004," *Cancer Causes Control,* vol. 22, pp. 1307-18, Sep 2011.

[3] C. Smigal, A. Jemal, E. Ward, V. Cokkinides, R. Smith, H. L. Howe, *et al.*, "Trends in breast cancer by race and ethnicity: Update 2006," *Ca-A Cancer Journal for Clinicians,* vol. 56, pp. 168-183, 2006.

[4] K. Unger-Saldaña, "Challenges to the early diagnosis and treatment of breast cancer in developing countries," *World journal of clinical oncology,* vol. 5, pp. 465-477, 2014.

[5] R. Radha, K. Batra, and S. Sekhar, "Classification of mammogram images: A survey," in *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*, 2018.

[6] V. Ajantha Devi and A. Nayyar, "Fusion of deep learning and image processing techniques for breast cancer diagnosis," in *Studies in Computational Intelligence* vol. 908, ed, 2021, pp. 1-25.

[7] D. Ponraj, M. Jenifer, P. Poongodi, and S. Manoharan, "A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer," *Journal of Emerging Trends in Computing and Information Sciences,* vol. 2, 01/01 2011.

[8] R. M. Rangayyan, F. J. Ayres, and J. E. Leo Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *Journal of the Franklin Institute,* vol. 344, pp. 312-348, 2007.

[9] P. National Council on Radiation, Q. C. o. R. P. Measurements Corporate Author, and Measurements, *Guide to Mammography and Other Breast Imaging Procedures*. Bethesda: NCRP Report No. 149, 2003.

[10] P. Sinthia and M. Malathi, "An effective two way classification of breast cancer images: A detailed review," *Asian Pacific Journal of Cancer Prevention,* vol. 19, pp. 3335-3339, 2018.

[11] S. B. Yengec Tasdemir, K. Tasdemir, and Z. Aydin, "A review of mammographic region of interest classification," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 10, 2020.

[12] R. Nithya and B. Santhi, "Computer aided diagnosis system for mammogram analysis: A survey," *Journal of Medical Imaging and Health Informatics,* vol. 5, pp. 653-674, 2015.

[13] H. Lee and Y. P. P. Chen, "Image based computer aided diagnosis system for cancer detection," *Expert Systems with Applications,* vol. 42, pp. 5356-5365, 2015.

[14] Y. Ali and S. Hamed, "Early breast cancer detection using mammogram images: A review of image processing techniques," *Biosciences Biotechnology Research Asia,* vol. 12, pp. 225-234, 2015.

[15] B. Mughal and M. Sharif, "Automated Detection of breast tumor in different imaging modalities: A review," *Current Medical Imaging Reviews,* vol. 13, pp. 121-139, 2017.

[16] L. M. Mina and N. A. M. Isa, "A review of computer-aided detection and diagnosis of breast cancer in digital mammography," *Journal of Medical Sciences (Faisalabad),* vol. 15, pp. 110-121, 2015.

[17] L. Schmarje, M. Santarossa, S. Schroder, and R. Koch, "A Survey on Semi-, Self-and Unsupervised Learning for Image Classification," *IEEE Access,* 2021.

[18] K. Ohri and M. Kumar, "Review on self-supervised image recognition using deep neural networks," *Knowledge-Based Systems,* vol. 224, 2021.

[19] A. Yates, R. Nogueira, and J. Lin, "Pretrained Transformers for Text Ranking: BERT and beyond," in *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 1154-1156.

[20] Y. N. Truong Vu, T. Tsue, J. Su, and S. Singh, "An improved mammography malignancy model with self-supervised learning," in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, 2021.

[21] A. R. Beeravolu, S. Azam, M. Jonkman, B. Shanmugam, K. Kannoorpatti, and A. Anwar, "Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN," *IEEE Access,* vol. 9, pp. 33438-33463, 2021.

[22] M. Ge, J. G. Mainprize, G. E. Mawdsley, and M. J. Yaffe, "Segmenting pectoralis muscle on digital mammograms by a Markov random field-maximum a posteriori model," *Journal of Medical Imaging,* vol. 1, 2014.

[23] F. Shih, *Image Processing and Pattern Recognition: Fundamentals and Techniques, Wiley-IEEE Press*, 2010.

[24] M. Ravikumar, P. G. Rachana, B. J. Shivaprasad, and D. S. Guru, "Enhancement of Mammogram Images Using CLAHE and Bilateral Filter Approaches," in *Cybernetics, Cognition and Machine Learning Applications*, Singapore, 2021, pp. 261-271.

[25] M. Biltawi, N. Al-Najdawi, and S. Tedmori, *Mammogram Enhancement And Segmentation Methods: Classification, Analysis, And Evaluation*, 2012.

[26] S. Srivastava, N. Sharma, S. K. Singh, and R. Srivastava, "A combined approach for the enhancement and segmentation of mammograms using modified fuzzy C-means method in wavelet domain," *Journal of medical physics,* vol. 39, pp. 169-183, 2014.

[27] S. J. Malebary and A. Hashmi, "Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram," *IEEE Access,* vol. 9, pp. 55312-55328, 2021.

[28] A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, E. R. E. Denton*, et al.*, "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis,* vol. 14, pp. 87-110, 2010/04/01/ 2010.

[29] R. Llobet, R. Paredes, and J. C. Pérez-Cortés, "Comparison of Feature Extraction Methods for Breast Cancer Detection," in *Pattern Recognition and Image Analysis*, Berlin, Heidelberg, 2005, pp. 495-502.

[30] B. Bektaş, İ. E. Emre, E. Kartal, and S. Gulsecen, "Classification of Mammography Images by Machine Learning Techniques," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 2018, pp. 580-585.

[31] X. Zhang, Y. Zhang, E. Y. Han, N. Jacobs, Q. Han, X. Wang*, et al.*, "Whole mammogram image classification with convolutional neural networks," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 700-704.

[32] S. A. Agnes, J. Anitha, S. I. A. Pandian, and J. D. Peter, "Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN)," *J Med Syst,* vol. 44, p. 30, Dec 14 2019.

[33] X. Li, Y. Hou, P. Wang, Z. Gao, M. Xu, and W. Li, "Transformer guided geometry model for flow-based unsupervised visual odometry," *Neural Computing and Applications,* vol. 33, pp. 8031-8042, 2021.

[34] J. B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya*, et al.*, "Bootstrap your own latent a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems*, 2020.

[35] T. Li, M. Feng, Y. Wang, and K. Xu, "Whole Slide Images Based Cervical Cancer Classification Using Self-supervised Learning and Multiple Instance Learning," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021*, 2021, pp. 192-195.

[36] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. Álvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *ArXiv,* vol. abs/2105.15203, 2021.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234-241.

[38] P. Bizopoulos, N. Vretos, and P. Daras, *Comprehensive Comparison of Deep Learning Models for Lung and COVID-19 Lesion Segmentation in CT scans*, 2020.