MULTIPLE INSTANCE LEARNING FOR HISTOPATHOLOGICAL BREAST CANCER IMAGE CLASSIFICATION

PROJECT REPORT

Submitted by

AJMI JALEEL

REG NO : TKM20CSCE01

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN

COMPUTER SCIENCE AND ENGINEERING

Under the guidance of
Prof. Thushara A

Thangal Kunju Musaliar College of Engineering
Kerala

JULY 2022

Thangal Kunju Musaliar College of Engineering
Dept. of Computer Science & Engineering

CERTIFICATE

This is to certify that this report titled ***Multiple Instance Learning for Histopathological Breast Cancer Image Classification*** is a bonafide record of the **Project** presented by **AJMI JALEEL (TKM20CSCE01)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Computer Science & Engineering** in **APJ Abdul Kalam Technological University** .

Coordinator	Supervisor	Head of the Department
Dr. Ansamma John	Prof. Thushara A	Dr. Dimple A Shajahan
Professor	Associate Professor	Associate Professor
Dept. of CSE	Dept. of CSE	Dept. of CSE
TKMCE	TKMCE	TKMCE

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of T.K.M College of Engineering for giving me an opportunity to present my project. I would like to thank **Dr. Dimple A Shajahan**, Associate Professor and Head of the Department, CSE, TKMCE, for her constant support and encouragement throughout the work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Prof. Thushara A**, Associate Professor, CSE, TKMCE, and project coordinator **Dr. Ansamma John**, Professor, CSE, TKMCE for their expert guidance, cooperation and immense encouragement. I also extend my thanks to the entire faculty members and staffs of the Department of Computer Science & Engineering, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents and friends, for their support and encouragement in the successful completion of this project work.

AJMI JALEEL

Abstract

Cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body. The availability of proper screening methods are important for detecting initial symptoms. Also, A tumor can be malignant (cancerous) or benign (not cancerous). A benign tumor is usually not a serious problem unless it presses on a nearby structure or causes other symptoms. A benign tumor can become quite large, but it will not invade nearby tissue or spread to other parts of your body. A malignant tumor can spread to other parts of your body. Cancer can be in many types and forms. breast cancer forms in breast cells and its very common type of cancer in women. The kind of breast cancer depends on which cells in the breast turn into cancer. The main factors that influence your risk include being a woman and getting older. Most breast cancers are found in women who are 50 years old or older. Need of best screening method is very important to identify benign and malignant tumors. Histopathological images have very important place for identifying breast cancer. A weakly supervised learning called multiple instance learning is used for the computer aided diagnosis of cancer. Without having to label every instance, multiple instance learning involves grouping instances (pictures) into bags (patients). more modern ones, such a deep learning-based approach and a non-parametric approach (MIL-CNN) is used here. The non-parametric technique, which is one of the MIL methods, offers the best overall results and, in some situations, enables the achievement of classification rates that are not possible with traditional (single instance) classification frameworks. The tests are performed on the publicly available BreaKHis dataset, which consists of 82 patients' microscopic biopsy images of 82 benign and malignant breast cancers. Above 95% accuracy lead to a better result.

Contents

List of Figures

Chapter 1

Introduction

Cancer as one of the most notorious diseases in the world, cancer has known to be one of the leading causes of death across the world. Cancer in any form is life-threatening and people often shy away from discussing it. In a recent study conducted in 2018, it was found that around 9.5 million people died that year owing to cancer. The World Health Organisation has revealed that cancer is the second leading cause of death across the world. The statistics in India are also no better and as per recent figures about 1300 people die every day owing to cancer of different types. Cancer types and causes have seen a steady increase in the past decade which does not bode well for the world population. Cancer types are usually named after the area they affect in the body - usually like skin, breast, lung, pancreas, blood, stomach among the others. However, in many cases, one single factor cannot be attributed or held responsible for causing cancer in an individual. The substances that are known to be cancer-causing or increasing the risks of cancer are known as carcinogens.

Breast cancer is the most common cancer in the UK. One in eight women will face it in their lifetime. Nearly 700,000 people living in the UK have experienced a diagnosis. Around 1,000 women in the UK die from breast cancer every month. That's one woman every 45 minutes whose life is cut short by the disease. Almost all these deaths are caused by secondary breast cancer, where cancer has spread from the breast to other parts of the body. Survival rates for breast cancer have thankfully increased dramatically. More than 80 percent of women who develop breast cancer will now live beyond five years. 40 years ago, it was only around half. But that doesn't mean we've got breast cancer under control. As women get older, they are at higher risk of breast cancer. So, its very important to have a better screening for this. The gold standard for the detection of breast cancer is histopathological imaging. For one patient, dozens of them are collected during the examination. Here, study the applicability of Multiple Instance Learning (MIL) for computer-aided diagnosis of breast cancer patients based on the analysis of histopathology pictures and offer a weakly supervised learning framework. Without having to label every instance, multiple instance learning involves grouping instances (pictures) into bags (patients).

Among the modern ones, such a deep learning-based approach and a non-parametric approach (MIL-CNN) is used here. Convolutional Neural Networks (CNN) have demonstrated its successful applications in computer vision, speech recognition, and natural language processing. The architecture of typical CNNs is a stack of convolutional, non-linear, pooling and fully-connected layers, followed by a loss function layer. It is designed to take advantages of local connections, shared weights, pooling and the use of many layers to learn highlevel representations of natural images, and it has demonstrated significant improvement over various benchmark object recognition datasets. Construct deep CNNs in a weakly supervised learning framework in order to lessen the impact of noisy training pairs. The tests are performed on the publicly available BreaKHis dataset, which consists of 82 patients' microscopic biopsy images of 82 benign and malignant breast cancers. The non-parametric technique, which is one of the MIL methods, offers the best overall results and, in some situations, enables the achievement of classification rates that are not possible with traditional (single instance) classification frameworks.

The remainder of this report is organized as follows. Chapter 2 recalls some related works used as reference for completing this study. Chapter 3 includes all background information used to carry out this study. Chapter 4 presents the dataset used for conducting the experiment and results of the study. Finally, conclusion and future works are provided in chapter 5.

Chapter 2

Related Works

Miao Sun, Tony X. Han.[1] Convolutional Neural Networks (CNN) have demonstrated its successful applications in computer vision, speech recognition, and natural language processing. For object recognition, CNNs might be limited by its strict label requirement and an implicit assumption that images are supposed to be targetobject- dominated for optimal solutions. However, the labeling procedure, necessitating laying out the locations of target objects, is very tedious, making high-quality large-scale dataset prohibitively expensive. Data augmentation schemes are widely used when deep networks suffer the insufficient training data problem. All the images produced through data augmentation share the same label, which may be problematic since not all data augmentation methods are label-preserving.Here, propose a weakly supervised CNN framework named Multiple Instance Learning Convolutional Neural Networks (MILCNN) to solve this problem. ReLU is used to replace traditional nonlinearity such as sigmoid and tangent functions, which tend to slow down learning procedure due to saturating effect. Multiple instance learning loss layer is developed in this paper to work harmony with data augmentation to fully explore the potential of training sets.

Alpaydin, E. , Cheplygina, V.[2] In multiple-instance (MI) classification, each input object or event is represented by a set of instances, named a bag, and it is the bag that carries a label. MI learning is used in different applications where data is formed in terms of such bags and where individual instances in a bag do not have a label. We review MI classification from the point of view of label information carried in the instances in a bag, that is, their sufficiency for classification. Our aim is to contrast MI with the standard approach of single-instance (SI) classification to determine when casting a problem in the MI framework is preferable. We compare instance-level classification, combination by noisy-or, and bag-level classification, using the support vector machine as the base classifier. We define a set of synthetic MI tasks at different complexities to benchmark different MI approaches. Our experiments on these and two real-world bioinformatics applications on gene expression and text categorization indicate that depending on the situation, a different decision mechanism, at the instance- or bag-level, may be appropriate. If the instances in a bag provide complementary information, a bag-level MI approach is useful; but sometimes the bag information carries no useful information at all and an instance-level SI classifier works equally well, or better.

Dietterich, T. G. , Lathrop, R. H[3] The multiple instance problem arises in tasks where the training examples are ambiguous: a single example object may have many alternative feature vectors (instances) that describe it, and yet only one of those feature vectors may be responsible for the observed classification of the object. This paper describes and compares three kinds of algorithms that learn axis-parallel rectangles to solve the multiple instance problem. Algorithms that ignore the multiple instance problem perform very poorly. An algorithm that directly confronts the multiple instance problem (by attempting to identify which feature vectors are responsible for the observed classifications) performs best, giving 89% correct predictions on a musk odor prediction task. The paper also illustrates the use of artificial data to debug and compare these algorithms.

I. Sutskever, O. Vinyals[4], Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector.

Wajahat Nawaz(B), Sagheer Ahmed [20],This work presents the details of a modified version of ALEXNET for classification of Breast Cancer slide images which was part of our entry to the ICIAR-2018 challenge on this topic. In this work, instead of designing our own CNN model, we fine tuned a good pre-trained model (ALEXNET) on breast cancer H E Stained Dataset. This approach includes convolutional, activation max-pooling layers; while reducing the size of fully connected layer in order to avoid overfitting.

Sema Nizam Abdulghani, Ahmed Freidoon Fadhil[24], This paper introduces an improved deep feature extraction CNN method for histapathological breast cancer image classification. The well-known Alexnet model were used to classifiy images from the BreakHis dataset using transfer learning. The BreaKHis dataset is preferred in the experimental works due to the huge number of sample images. Two different models were considered and performed. In the first model, the pre-trained Alexnet that uses six patches from each of the images of the dataset was conducted. While the second model is based on pre-trained Alexnet with SVM classifier.

Maron Lozano-Pérez [26],Diverse density is closely related to the idea of the APR. The DD defines a function over the feature space, such that it is high at points that are both close to instances from positive bags, and far away from instances which are in neg- ative bags. The DD algorithm attempts to find the local maxima of this function (called the positive instance targets or prototypes) by maximizing diverse density (i.e. conditional likelihood) over the instance space, using gradient ascent with multiple starting points.

Wang Zucker[29],The Citation-kNN, an adaptation of k-nearest neighbors (k- NN) algorithm, is the first non-parametric approach. The principle is to first apply the k -NN algorithm to bags, where the distance between bags is measured with the minimum Hausdorffdistance.

Venkatesan et al[7] This recent technique is designed as a modified version of the k -NN classifier. The non-parametric MIL approach employs a new formulation based on distances to k- nearest neighbors. The idea is to parse the MIL feature space with a Parzen window technique, using different sized regions. Con- versely to the majority vote used in k -NN, the vote contributions are the kernelized distances in the feature space. Non-parametric MIL has shown enhanced robustness to labeling noise on various datasets.

Chapter 3

Methodology

Weakly supervised learning can be handled elegantly with the help of multiple instance learning (MIL). The advantage of the MIL paradigm is that instances are naturally grouped in labelled bags without the need for all the instances of each bag to have individual labels, in contrast to strong supervised learning where every training instance is given a discrete or real-valued label.

3.1 Multiple Instance Learning (MIL)

A prediction function is inferred from a series of labelled training examples in the machine learning discipline known as supervised learning in order to map each input occurrence to its associated label. In a typical setup, cases with their corresponding labels are part of the training dataset. The fundamental barrier to having large-scale datasets is the expensive data-labeling process using human-based ground-truth descriptions, even when cases are relatively inexpensive to get. Due to this problem, machine learning has developed a new paradigm known as weakly supervised learning, which is used when training data is only partially labelled. Weakly supervised learning can be handled elegantly with the help of multiple instance learning (MIL). A bag is labelled positive in the binary classification situation if it has at least one positive instance, and it is labelled negative if all of its examples are negative. MIL algorithms attempt to categorise either unseen occurrences or unseen bags using this training data organised in labelled bags. Although MIL has various uses in medical imaging, as demonstrated in a recent study, there is increased interest in its use, particularly for the categorization of histopathological images. The gold standard for cancer diagnosis is histopathological pictures, which are microscopic photographs of tissue for disease evaluation.

3.2 Multiple Instance Learning Convolutional Neural network (MIL-CNN)

CNNs (Convolutional Neural Networks) have shown its effective uses in speech recognition and computer vision processing natural language and recognition. To identify objects, CNNs could be constrained by the stringent labelling requirements and a tacit understanding that targets are supposed to be pictures dominated in search of ideal answers.CNNs (Convolutional Neural Networks) have shown its effective uses in speech recognition and computer vision processing natural language and recognition. To identify objects, CNNs could be constrained by the stringent labelling requirements and a tacit understanding that targets are supposed to be pictures dominated in search of ideal answers.

Deep learning algorithms are transforming numerous industries. tasks using artificial intelligence, such as language processing Speech recognition, computer, and eyesight. Deep learning algorithm's achievements are the outcome of its outstanding capacity to find complex high-dimensional data structures with minimal manual intervention engineering. The ImageNet Challenge's innovations has shown that effective feature representations be automatically taught from data, progressing from conventional methods based on features created by hand. The deep learning algorithm with the greatest degree of success Convolutional Neural Networks are used for image recognition. (CNN). The standard CNN's architecture is a stack of convolutional, non-linear, completely connected, and pooling layers, then a layer with a loss function.

The impact of noisy training pairs was lessened by deep CNNs in a framework for poorly supervised learning. Rather than giving all of the generated photos labels we treat the created images as a bag and the original as seen in label as a bag-level label. This occurrence is logically the Multiple Instance Learning problem (MIL). A bag is marked as positive for binary MIL if it contains a minimum of one good example, and it is deemed negative if there are none. Its occurrences are detrimental. MIL should therefore be incorporated into The promise of the deep learning algorithm would be completely realised. training goals and improve performance

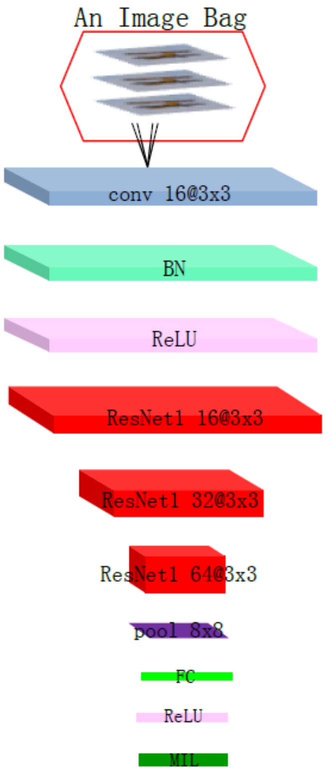


Figure 3.1: MIL-CNN

3.3 Formulation of MIL- CNN

For multiple instance learning, training instances are not singletons. Instead, they come in "bags", where all the examples in a bag share the same label. For object recognition, regions in each image are considered as a bag.

$\mathrm{B}=\left\{\mathrm{x}^1,\mathrm{x}^2\ldots\mathrm{x}^m\right\}, \mathrm{x}^m$ is m_{th} region in the image. The loss function w.r.t to the bag B is

$$f_{\text{loss}} = -\sum_{i=1}^{C} y_i \log\left(p\left(c_i = 1 \mid B\right)\right)$$

(3.1)

According to concept of multiple instance learning, B is a negative bag for the i_th category if all the instances in the bag are negative:

$$p\left(c_i = 0 \mid B\right) = \Pi_{j=1}^{m}\left(1 - p\left(c_i = 1 \mid \mathbf{x}^j\right)\right) \tag{3.2}$$

The gradients will be:

$$\frac{\partial f_{loss}}{\partial h_i^j} = \lambda\left(1 - y_i\right) \tag{3.3}$$

3.4 Transfer Learning Using Alexnet

Training a deep convolutional neural network from scratch requires massive amount of data and significant computational power. However, to collect a large amount of data in medical field is costly and difficult, but this can be solved by some clever tricks such as mirroring, rotating and fine tuning pre-trained neural networks. Therefore, fine tune a deep convolutional neural network (ALEXNET) by changing and inserting input layer convolutional layers and fully connected layer. One of the problems associated with convolution neural network parameter learning is that it requires massive amounts of data and significant computing power. The computational problem can be overcome by employing GPU computing. However, collecting very large sized medical imaging datasets can be difficult and time-consuming. This problem, however, can be solved to some extent by some clever tricks such as mirroring, rotating images, extracting overlapping patches and fine tuning(Transfer learning). Transfer learning basically takes an existing pre-trained deep neural network such as ALEXNET or GOOGLE-NET and fine tunes its parameters by changing the last three layers based on the application specific dataset.

It consists of multiple sets of convolutional layers, pooling layers and fully-connected layers stacked on top of each other. The job of the convolutional layers is to extract local features in the input images. A convolutional layer is typically followed by a pooling layer that reduces computational complexity by performing non-linear down-sampling. Pooling layers also introduce translation invariance into the network. In the upper and final layers, all the units are generally connected to each other and thus are termed fully-connected layers. Initially, it is trained on “Imagenet” dataset which contains 1000 distinct classes consisting of

1.2 million labelled images. AlexNet has been trained on approximately 1.2 million images from the ImageNet Dataset. The Alexnet structure has eight layers for classification of 1000 different objects. For this reason, the model has gained a rich feature representation to classify images from different applications. The first five layers, the convolutional layers, from the pre-trained Alexnet are saved as fixed feature extractors, while the last three layers, fully connected layers, were replaced with new set of layers that can classify two classes only. The network parameters will be updated by using the Stochastic Gradient Descent (SGD) method with backpropagation. A complete pass of the algorithm over the entire training set is called an epoch. The mini-batch size is the subset of the training dataset used by the SGD to update network parameters. In contrast, the rate of adjusting the weights of the network to the gradient is called the learning rate.

The experimental results are presented in the following chapter.

Chapter 4

Experimental Results and Discussions

4.1 Experimental Data

The experiments are conducted on the public BreaKHis dataset which contains about 80 0 0 microscopic biopsy images of benign and malignant breast tumors, originating from 82 patients. Among the MIL methods the non-parametric approach has the best overall results, and in some cases allows to obtain classification rates never reached by conventional (single instance) classification frame- works.

SOB_B_A-14-225 49AB-40-001

SOB_B_A-14-225 49AB-40-002

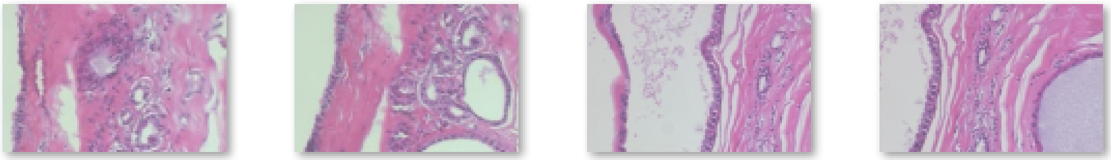
SOB_B_A-14-225 49AB-40-003

SOB_B_A-14-225 49AB-40-004

Figure 4.1: Benign Tumor Histopathological Image

A benign tumor is an abnormal but noncancerous collection of cells. It can form anywhere on or in your body when cells multiply more than they should or don't die when they should. A benign tumor is not malignant. It grows more slowly, has even borders and doesn't spread to other parts of your body.

Malignant tumors have cells that grow uncontrollably and spread locally andor to distant sites. Malignant tumors are cancerous (ie, they invade other sites). They spread to distant sites via the bloodstream or the lymphatic system. This spread is called metastasis. Metastasis can



SOB_B_A-14-225 49CD-100-001

SOB_B_A-14-225 49CD-100-002

SOB_B_A-14-225 49CD-100-003

SOB_B_A-14-225 49CD-100-004

Figure 4.2: Malignant Tumor Histopathological Image

occur anywhere in the body and most commonly is found in the liver, lungs, brain, and bone. Malignant tumors can spread rapidly and require treatment to avoid spread.

4.2 Settings

Images were acquired in RGB color space, with a resolution of 752 ×582 using magnifying factors of 40 ×, 100 ×, 200 ×and 400 ×. This image is acquired from a single slide of breast tissue containing a malignant tumor (breast cancer). The highlighted rectangle (manu- ally added for illustrative purposes only) is the area of interest selected by the pathologist to be detailed in the next higher magnification. To date, the database is composed of 7909 images divided into benign and malignant tumors. The BreaKHis dataset has been ran- domly divided into a training set (70%) and a testing set (30%). A size of 64 ×64 pixel was chosen for the patch size, as it has been shown to be particularly relevant for CNN-based classification. For training, 1000 patches are randomly extracted from each input image. For test, to preserve computa- tional cost, a grid of non-overlapping patches is extracted, yield- ing around 100 patches per image.

4.3 Result Analysis

Here, its analysed that the given data is malignant or benign by using the methods MIL-CNN and HistoBreast net.

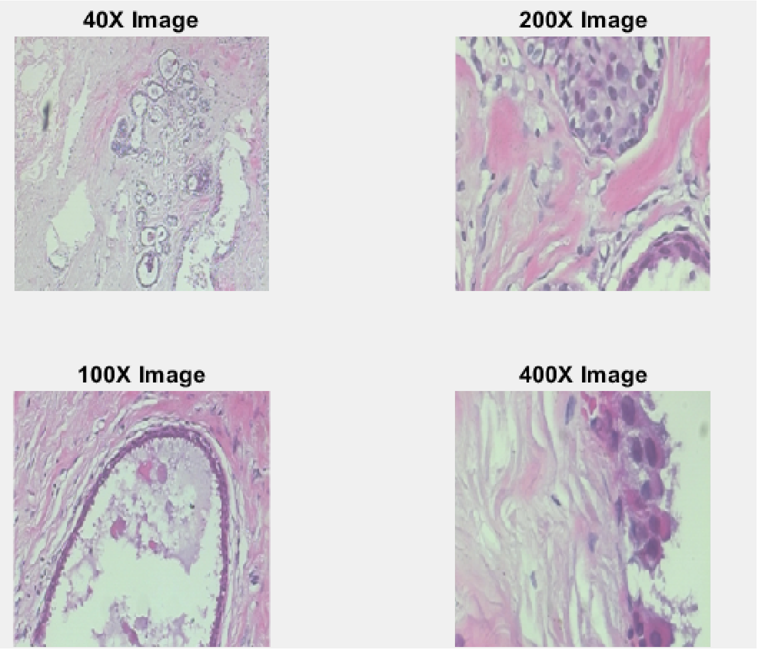


Figure 4.3: Benign

4.4 Performance Comparison

To evaluate the effectiveness of the proposed MIL-CNN and the transfer learning method, MIL-CNN have above 95% of accuracy and also the error rate is less. It requires more data to be used in comparison with Histo-breast net. It is not to be considered as a unique model for identifying malignant as well as benign tumors. HistoBreast Net is a model can be represented as a uniquee model in terms of accuracy as well as it can be used as substitute to other processed data. That means, can be used as a balance of a processed classifier. It gives above 98% accuracy and less error rate.

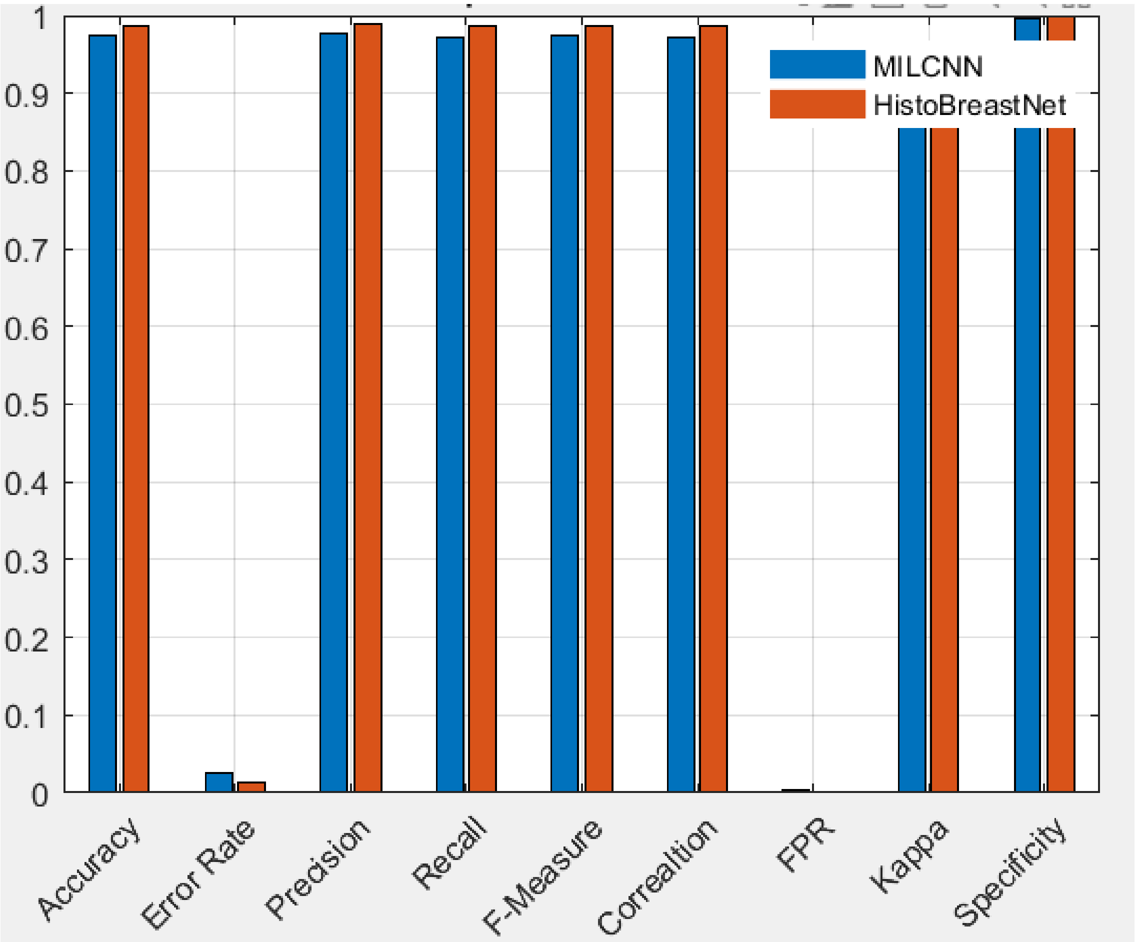


Figure 4.4: Comparison

Chapter 5

Conclusion and Future Work

Multiple instance learning provides a classification framework that is particularly adapted to computer-aided diagnosis based on histopathological image analysis. In the case of the BreaKHis dataset, several hundreds of images are available per patient. The patient can thus be considered as a bag, which is labeled with its diagnosis. MILCNN are particularly efficient for the tasks of patient and image classification. Patient classification rates can reach up to 95.5%, a level never reached by conventional classification frameworks, which enhances the fact that instances are complementary and can be fruit- fully considered in a MIL framework. MIL can thus leverage digital histopathological image classification and analysis to improve computer-aided diagnosis, without the need to label all the images.

Deep feature extraction can be considered a type of transfer learning. The activation layers of the Alexnet model can be used to extract the feature vectors instead of fine-tuning the pre-trained alexnet model. The earlier layers from the pre-trainedAlexnet model represents low level features from images such as edges, while the last layers, fully connected layers, represents higher level features for image classification. Features from the first two layers of the fully connected layers were frequantly used in classification problems. It acquires more that 98% of accuracy.

Breast cancer is one of the leading causes of women death worldwide currently. Developing a computer-aided diagnosis system for breast cancer is very relevant in this scenario. Imaging technologies gained a huge improvement for BC diagnosis, the final decesion in BC diagnosis depends on the visual inspection of the pathologists for samples under the microscope.

References

[1] Sun, M. , Han, T. X. , Liu, M.-C. , Khodayari-Rostamabad“Multiple instance learning convolutional neural networks for object recognition” In International conference on pattern recognition (ICPR) (pp. 3270–3275)

[2] Alpaydin, E. , Cheplygina, V. , Loog, M. , Tax, D. M. J., “Single-vs. multiple-in- stance classification. Pattern Recognition,” 48 (9), 2831–2838 .

[3] Dietterich, T. G. , Lathrop, R. H. , Lozano-Prez, T., “Solving the multiple in- stance problem with axis-parallel rectangles,” Artificial. Intelligence, 89 (1), 31–71.

[4] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in arXiv:1409.3215, 2014.

[5] ang, J. , Zucker, J.-D. (20 0 0). ”Solving the multiple-instance problem: A lazy learn- ing approach”, In Proceedings of the seventeenth international conference on ma- chine learning . In ICML (pp. 1119–1126) . .

[6] Andrews, S. , Tsochantaridis, I. , Hofmann, T. (2002). ”Support vector machines for multiple-instance learning”, In Proceedings of the 15th international conference on neural information processing systems . In NIPS’02 (pp. 577–584). Cambridge, MA, USA: MIT Press .

[7] Venkatesan, R. , Chandakkar, P. S. , Li, B. (2015), ”Simpler non-parametric methods provide as good or better results to multiple-instance learning”, In 2015 ieee in- ternational conference on computer vision (ICCV) (pp. 2605–2613) .

[8] Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan, “Cnn: Single-label to multi-label,” in arXiv:1406.5726, 2014.

[9] M. Sun, T. X. Han, and X. Xu, “Latent model ensemble with autolocalization,” in arXiv:1604.04333, 2016.

[10] M. Malinowski and M. Fritz., “Learnable pooling regions for image classification.” in arXiv:1301.3516, 2013.

[11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in arXiv:1409.1556, 2015.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” in arXiv:1409.0575, 2014.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Neural Information Processing Systems, NIPS 2012.

[14] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013 vol. 35, no. 8, pp. 1915-1929.

[15] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, “Joint training of a convolutional network and a graphical model for human pose estimation,” in arXiv:1406.2984, 2014.

[16] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, “Joint training of a convolutional network and a graphical model for human pose estimation,” in arXiv:1406.2984, 2014.

[17] Zhang, B.: ”Breast cancer diagnosis from biopsy images by serial fusion of Random Subspace ensembles”, In: 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), vol. 1. IEEE (2011)

[18] Krizhevsky, A., Ilya S., Hinton, G.E, ”Imagenet classification with deep convolutional neural networks”, In Advances in Neural Information Processing Systems (2012)

[19] Szegedy, C, ”Going deeper with convolutions”, In Proceedings of the IEEE conference on computer vision and pattern recognition (2015)

[20] Litjens, G., ”Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis”, Sci. Rep. 6, 26286 (2016)

[21] Ara´ujo, T, "Classification of breast cancer histology images using Convolutional Neural Networks", PloS one 12(6), e0177544 (2017)

[22] Macenko, M, "A method for normalizing histology slides for quantitative analysis", In 2009 IEEE International Symposium on Biomedical Imaging, ISBI2009, From Nano to Macro. IEEE (2009)

[23] Brook, A., El-Yaniv, R., Issler, E., Kimmel, R., Meir, R., Peleg, D, "Breast cancer diagnosis from biopsy images using generic features and SVMs", pp. 1–16 (2007)

[24] Deniz, E., et al. (2018). "Transfer learning based histopathologic image classification for breast cancer detection."

[25] Niu, X.-X. and C. Y. J. P. R. Suen (2012). "A novel hybrid CNN–SVM classifier for recognizing handwritten digits." 45(4): 1318-1325.

[26] Xu, Y. , Zhu, J.-Y. , Eric, I. , Chang, C. , Lai, M. , Tu, Z. (2014), "Weakly supervised histopathology cancer image segmentation and classification", Medical Image Analysis, 18 (3), 591–604 .

[27] Zhang, Q. , Goldman, S. A. (2001). "EM-DD: An improved multiple-instance learning technique", In advances in neural information processing systems (pp. 1073–1080). MIT Press .

[28] Zhou, Z.-H. (2017), "A brief introduction to weakly supervised learning", National Sci- ence Review, 00 , 1–10 .

[29] Mercan, C. , Aksoy, S. , Mercan, E. , Shapiro, L. G. , Weaver, D. L. , Elmore, J. G. (2018), "Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images", IEEE Transactions on Medical Imaging, 37 (1), 316–325 .

[30] Pathak, D. , Shelhamer, E. , Long, J. , Darrell, T. (2014), "Fully convolutional multi-class multiple instance learning", CoRR, abs/1412.7144 .