Deep Learning for Digital Pathology Using Representation Learning

Arockia Liborious[1], Anish Anurag[2], Govind Kumar Jha[3]

[1] Principal Consultant, Clariant, Mumbai, Maharashtra, India, arockialiborious@gmail.com, 7875552111
[2] Vinoba Bhave University, Hazaribag, Jharkhand, India, anish.jnu08@gmail.com, 7004095596
[3] Vinoba Bhave University, Hazaribag, Jharkhand, India, jhagovi@gmail.com, 9798315444

Abstract

Disease diagnosis if made early and accurately will save lives. Pathologists manually diagnose diseases based on tissue samples. The diagnostic process is usually time-consuming and costly. As a result, automated tissue sample analysis from histopathology images is important for early diagnosis and treatment. The collection, management, exchange, and interpretation of pathology knowledge including slides and data in a digital environment is referred to as digital pathology. Digital pathology is a field that combines pathology and computers, and it has the potential to replace traditional microscope-based diagnosis soon. In this paper, a deep learning-based representation learning method for automatically classifying histopathological images is proposed. Two well-known and current pre-trained convolutional neural network (CNN) models, VGG-16, and Inception-v3, have been used for feature extraction. The VGG-16 and Inception-v3 pre-trained model were tested on the color images of the Kimia Path24 dataset. Additionally, to reduce the number of extracted features, a Principal Component Analysis (PCA) was done keeping in mind the future applicability of this proposed model in real life scenarios. According to the obtained results, it may be said that the proposed pre-trained models can be used for fast and accurate classification of histopathology images and assist pathologists in identifying some of the most critical and life-threatening diseases like cancer at an early stage and improve the chances of survival.

Keywords: Deep Learning, Classification, Digital Pathology, Transfer Learning, Representation Learning, Convolution Neural Network

1. Introduction

The integration of algorithms for classification and retrieval in medical images through effective machine learning schemes is at the forefront of modern medicine [1]. These tasks are crucial as they are instrumental in detecting and analysing abnormalities and

malignancies faster and allow well informed decision making on time. Digital pathology is an emerging area in the field of clinical diagnosis where faster decision making with the help of advanced machine learning algorithms can save lives [2]. As an age-old method of analysing tissue samples, the process of archiving microscopic information of specimens has been primarily achieved through the use and storing of glass slides [3]. Glass slides are not only fragile in nature, but hospitals and clinics need large and specially prepared storage rooms to store specimens. To store these specimens, special infrastructural enhancements need to be made so that the specimens are not destroyed.

One of the most promising ways of overcoming the drawback of glass slides, is the use of Whole Slide imaging (WSI), also commonly known as digital pathology. WSI not only can preserve the quality of the image and prevent it from decaying but also has multiple other benefits [3, 4]. One of the most important aspects of digital pathology is it allows multiple experts to read, analyse the same images at the same time, the whole process of studying the images can happen parallel between experts instead of following a sequential step which kills a lot of time and delays the process of disease identification.

Digital Pathology involves investigation of a biopsy or a surgical tissue specimen at microscopic level. These tissue samples are chemically processed and then sectioned on a glass slide for the study and analysis of cellular morphology mostly for cancer diagnosis. One of the steps involved in tissue segmentation so that it can viewed under a microscope is to dye the tissue. One of more stains of Haematoxylin-Eosin and Immunohistochemical (IHC) is used for this purpose. The nuclei regions are stained in dark blue colour by Haematoxylin and the other structures like cytoplasm, stroma etc., are stained with pink colour. IHC is used to determine the cancer stage whether it is benign or malignant based on the presence or absence of proteins. After the process of staining, digital images are generated using fast slide scanners which contain one or multiple lenses to magnify the images at X20 or X40 magnification. Uniform light spectrum is used to illuminate the tissue slide. The slide scanners are provided with standard packages for corrections in spectral and spatial illumination variation. Lymphocyte is the white blood cell which plays major role in immune system of the body. Epithelial tissues line the outer surfaces of organs, blood vessels and inner surfaces of cavities of human body. Lymphocyte Nuclei (LN) have regular shape and are smaller in size than Epithelial Nuclei (EN). EN's in high grade cancer tissues are larger in size and have clearly visible nucleoli. Also, they show heterogeneous chromatin distribution and irregular boundaries called nuclear pleomorphism. The problems associated with detection, segmentation and classification of nuclei are due to variation in slide preparation, image acquisition like artifacts caused during image compression, noise etc., and the overlapping clusters of nuclei. The aspect of nuclei plays a major role in evaluating the existence of cancer and its severity.

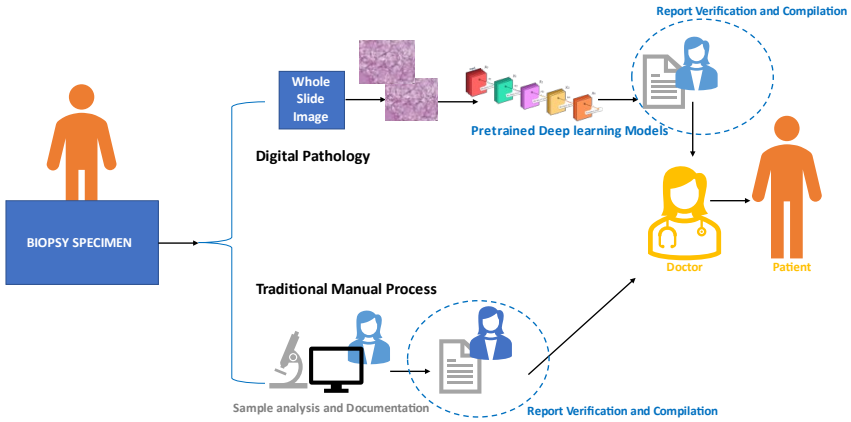

Figure 1. Proposed Framework for Histopathological Image Classification

In 1999, Wetzel and Gilbertson developed the first auto- mated WSI system [5], utilizing high resolution to enable pathologists to buffer through immaculate details presented through digitized pathology slides. Ever since, pathology bounded by WSI systems is emerging into an era of digital specialty, providing solutions for centralizing diagnostic solutions by improving the quality of diagnosis, patient safety, and economic concerns [6]. For larger size images, the parameters to be estimated, required computational power and memory also increase. Hence, the images must be resized to smaller images which results in loss of information at cellular level and there will be decrease of identification accuracy. Therefore, the entire Whole Slide Image (WSI) is divided into partial regions called patches and each patch is analysed independently. For increased patch sizes, the accuracy [7].

II. Literature Survey

The broad dimensionality of the image in digital pathology makes computation and storage difficult; thus, contextually understanding regions of interest in an image aids in faster diagnosis and identification by using soft-computing techniques [8]. Cell structures such as cell nuclei, glands, and lymphocytes are observed to have prominent characteristics that serve as a hallmark for detecting cancerous cells, particularly in histopathology. Researchers conclude that by correlating histological trends with protein and gene expression, conducting exploratory histopathology image analysis, and performing computer assisted diagnostics (CADx), pathologists will be able to make better decisions [9]. In recent years, WSI technology has been steadily establishing laboratory standards as a method of digitizing pathology slides for more effective diagnostic, educational, and research purposes [10]. The author of this paper [11] proposed the method of picture foreground extraction using a graph cuts based binarization, resulting in enhanced automatic detection and segmentation of cell nuclei. Multiple scale Laplacian of Gaussian filtering was used to detect nuclear seed points, and a graph cuts-based algorithm is used for segmentation. Overall, the process achieved an accuracy of more than 86%. The authors

of this paper [12] contrasted the classification of histopathological images using Local Binary Patterns (LBP), Bag-of-Visual (BOV) terms, and deep features. They presented a new dataset, KIMIA Path960, which contains 960 histopathological images from 20 different tissue types and demonstrated a deep feature recognition accuracy of 94.72%. Similar deep learning techniques for histopathological image classification and fusion of features when dataset is small to improve model accuracy is discussed in these papers [13] [13][14]. Mitosis identification is critical for cancer grading and is one of the most significant factors in cancer prognosis. Authors of this paper [15] proposed a supervised model for detecting mitosis signature from histopathological images using deep learning. It includes five convolution layers, four max-pooling layers, four rectified linear units (ReLU), and two fully connected layers. Features such as morphological, textural, and intensity are all handcrafted and are incorporated. The method provides an efficient second opinion for breast cancer grading from whole slide images.
This paper found a large gap in faster classification and higher precision with restricted training image data. This substantial gap has been closed by representation learning. The findings are positive, and so far, they have shown improved classification.

III. Materials and methods

Image segmentation is the problem of outlining relevant objects in images, the intent is to use the representation learning feature using ImageNet pretrained models to auto decide the feature selection that not only improves the model accuracy but also improves the model runtime. Representation learning is a technique that allows the algorithm to discover the representations needed for a feature detection or classification of raw data. The data set used here is KIMIA Path960 which contain 960(=20×48) images of manually selected 48 regions of interest. The images were captured by Tissue Scope LE 1.0. The scans were done with a 0.75 NA lens in the bright region.

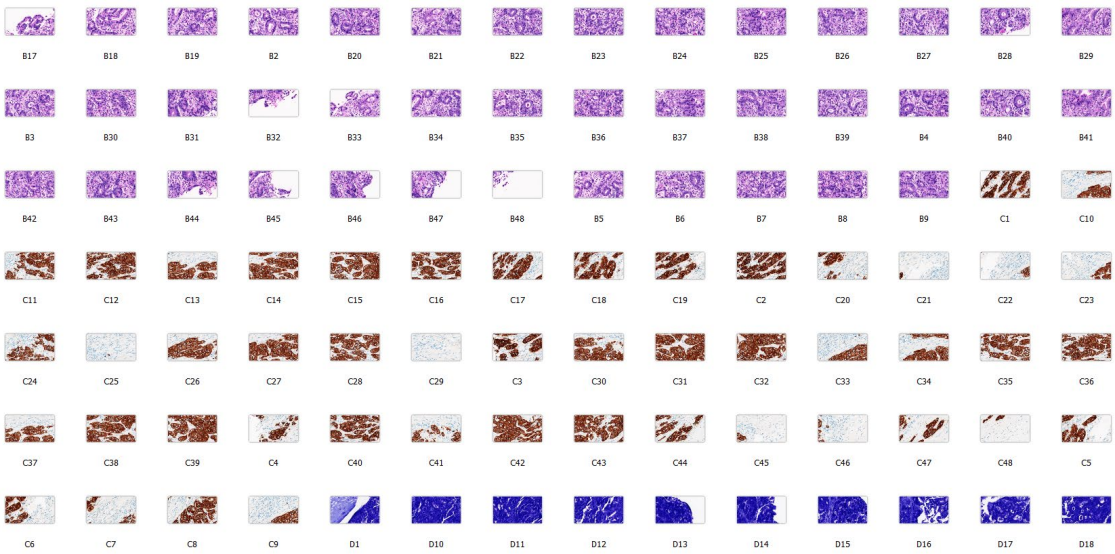

Figure 2. KIMIA Path960 Dataset – Sample View

III A. Representation learning models

To provide preliminary results for setting a benchmark line for the proposed data set, we performed two sets of feature extraction experiments using the pretrained models of VGG-16 and Inception-v3. Following the feature selection process, a multilayer perceptron and SVM classifiers were used to compare results. To understand the applicability in real life scenario where the image numbers will be higher in number, prevent overfitting and to better generalize the model by producing independent, uncorrelated features PCA was applied on both VGG-16 and Inception-v3 features. After PCA a multilayer perceptron and SVM classifiers were used to compare results. All the models were evaluated using 10-fold cross validation technique as the dataset size is small and lacks variety of tissue images.

In all below experiments two classifiers namely neural network and SVM were used.

- Experiment 1: Auto extracted features using VGG16 pre trained model was used.
- Experiment 2: Auto extracted features using Inception-v3 pre trained model was used.
- Experiment 3: PCA was performed on the auto extracted features of VGG16.
- Experiment 4: PCA was performed on the Auto extracted features of Inception-v3.

IV. Results and discussion

In this study two pretrained models namely VGG16 and Inception-v3 (which were trained on ImageNet dataset) were used to auto extract the features of Kimia Path960 dataset. All model results were computed using 10-fold cross validation.

Feature Extraction	Number of Features
VGG16	4096
Inception v3	2048
PCA on VGG16 Features	150
PCA on Inception v3 Features	150

Table 1. Details of feature dimension

IV A. Experiment 1 (VGG16 auto features)

VGG16 auto extracted features were used to classify the images via neural network and SVM classifier. Both classifiers performed the same in terms of accuracy and F1 score.

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

Where, TP - True Positives, TN - True Negatives, FP - False Positives, FN - False Negatives

$$F1\ Score = \frac{(2*(Recall * Precision))}{(Recall + Precision)} \quad (2)$$

Where, Precision = TP/TP+FP, Recall = TP/TP+FN

Table 2 summarises the results of this experiment and supporting the claim that representation learning technique can yield improved classification accuracy.

Evaluation Metric	Neural Network	SVM
AUC	0.999	0.999
Accuracy	0.954	0.953
F1 Score	0.954	0.953
Precision	0.955	0.955
Recall	0.954	0.953

Table 2. Model Summary for VGG16 Feature Extraction

IV B. Experiment 2 (Inception v3 auto features)

Inception-v3 auto extracted features were used to classify the images via neural network and SVM classifier. Both classifiers gave almost similar results in this case and the same can be seen in Table 3.

Evaluation Metric	Neural Network	SVM
AUC	0.999	0.999
Accuracy	0.955	0.947
FI Score	0.955	0.947
Precision	0.956	0.948
Recall	0.955	0.947

Table 3. Model Summary for Inception v3 Feature Extraction

The model run time for experiments 1 and 2 were recorded and listed in table 4. Inception-v3 based models run ~80% faster than VGG16 based models.

Features	Classifier	Accuracy (%)	Number of Features	Run Time (mins)
VGG16	Neural Network	0.954	4096	22.7
VGG16	SVM	0.953	4096	22.3
Inception v3	Neural Network	0.955	2048	4.4
Inception v3	SVM	0.947	2048	3.5

Table 4. Model Runtime Summary for Auto Extracted Features

IV C. Experiment 3 (PCA on VGG16 auto features)

PCA was applied on VGG16 extracted features and from figure 3 it is evident that 150 Principal Components (PC) were able to explain more than 80% variance. The model summary is presented in table 5 where SVM classifier gives better accuracy.

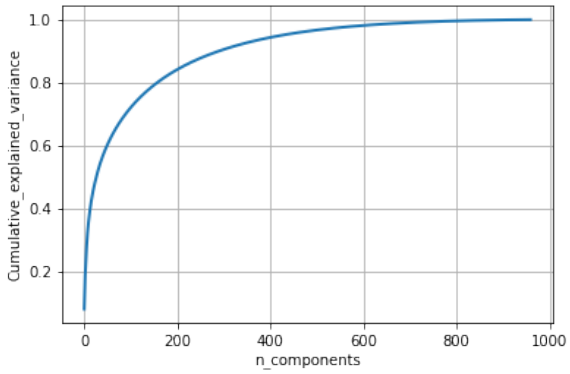

Figure 3: Explained Variance vs Number of Principal Components (VGG16)

Evaluation Metric	Neural Network	SVM
AUC	0.985	0.994
Accuracy	0.863	0.903
FI Score	0.856	0.904
Precision	0.865	0.909
Recall	0.863	0.903

Table 5. Model Summary for PCs of VGG16 Extracted Features

IV D. Experiment 4 (PCA on Inception v3 auto features)

PCA was performed on Inception-v3 auto extracted features and like experiment 3, 150 PCs explained more than 80% variance.

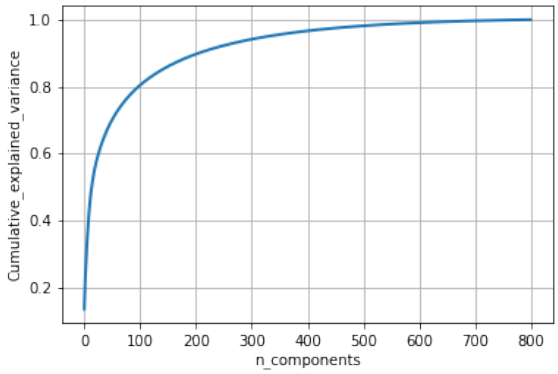

Figure 4: Explained Variance vs Number of Principal Components (Inception-v3)

The model results were satisfactory, and PC's give better generalization which in turn helps prevent the model overfitting issue.

Evaluation Metric	Neural Network	SVM
AUC	0.996	0.999
Accuracy	0.906	0.948
FI	0.906	0.948
Precision	0.909	0.95
Recall	0.906	0.948

Table 6. Model Summary for PCs of Inception v3 Extracted Features.

The model run time for experiments 3 and 4 were recorded and listed in table 7. PCs of Inception-v3 based models run ~87% faster than PCs of VGG16 based models. SVM classifier on PCs of Inception-v3 auto features performs better in terms of both time and accuracy. SVM are known to perform well when the number of features is less than the number of samples. ANNs can overfit if training samples are less - a problem that SVMs do not have. Hence SVM seem to perform better and also better generalizes the feature vector and thus the model.

Features	Classifier	Accuracy (%)	Number of Features	Run Time (mins)
PCA (VGG16)	Neural Network	0.863	150	18.5
PCA (VGG16)	SVM	0.903	150	18.3
PCA (Inception v3)	Neural Network	0.906	150	2.5
PCA (Inception v3)	SVM	0.948	150	2

Table 7. Model Runtime Summary for PCA

V. Conclusions and future scope

As seen the above experiments, without any use of image preprocessing techniques and handcrafted feature selection, the Auto Feature selection process using pretrained models provide a higher level of accuracy. Also, when the dimensionality analysis is performed, the model runtime reduces further and the model accuracy with respect to Inception V3.0 remains above 90%. As a part of future scope of study, an appropriate image preprocessing technique can be evaluated which would enhance and blur certain parts of the WSI image based on the intensity of light. One of the experiments that can be explored along with the pretrained models is Trained Weka Segmentation technique as an image preprocessing step. This would not only reduce the time consumed in identifying the right handcrafted feature selection process but also positively impact the overall accuracy of the model. As a part of future scope, the efficacy of the preprocessing technique and pretrained models can be tested and a time comparison study can be performed by benchmarking the time

spent in manually analyzing the glass slides by pathologists. As seen from the previous studies a considerable amount of time is spent in both manual as well as previous suggested machine learning models in data preparation and preprocessing steps. The future studies can focus on reducing the time spent and further improve the accuracy and model generalization for real time data.

References

[1] Carpenter AE, Kamentsky L, Eliceiri KW (2012) A call for bioimaging software usability. Nat Methods 9:666–670. https://doi.org/10.1038/nmeth.2073

[2] Madabhushi A, Lee G (2016) Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 33:170–175. https://doi.org/10.1016/j.media.2016.06.037

[3] Saco A, Bombi JA, Garcia A, et al (2016) Current Status of Whole-Slide Imaging in Education. Pathobiology 83:79–88. https://doi.org/10.1159/000442391

[4] Gabril MY, Yousef GM (2010) Informatics for practicing anatomical pathologists: Marking a new era in pathology practice. Mod Pathol 23:349–358. https://doi.org/10.1038/modpathol.2009.190

[5] Ho J, Parwani A V., Jukic DM, et al (2006) Use of whole slide imaging in surgical pathology quality assurance: Design and pilot validation studies. Hum Pathol 37:322–331. https://doi.org/10.1016/j.humpath.2005.11.005

[6] Ghaznavi F, Evans A, Madabhushi A, Feldman M (2013) Digital imaging in pathology: Whole-slide imaging and beyond. Annu Rev Pathol Mech Dis 8:331–359. https://doi.org/10.1146/annurev-pathol-011811-120902

[7] Komura D, Ishikawa S (2017) Machine learning methods for histopathological image analysis. arXiv

[8] Shirazi SH, Naz S, Razzak MI, et al (2018) Automated pathology image analysis. Elsevier Inc.

[9] Caicedo JC, González FA, Romero E (2011) Content-based histopathology image retrieval using a kernel-based semantic annotation framework. J Biomed Inform 44:519–528. https://doi.org/10.1016/j.jbi.2011.01.011

[10] Pantanowitz L, Evans A, Pfeifer J, et al (2011) Review of the current state of whole slide imaging in pathology. J Pathol Inform 2:36. https://doi.org/10.4103/2153-3539.83746

[11] Al-Kofahi Y, Lassoued W, Lee W, Roysam B (2010) Improved automatic detection and segmentation of cell nuclei in histopathology images. IEEE Trans Biomed Eng 57:841–852. https://doi.org/10.1109/TBME.2009.2035102

[12] Prakash RM, Shantha Selva Kumari R. (2018) Histopathological Image Analysis in Medical Decision Making. 139–153. https://doi.org/10.4018/978-1-5225-6316-7.ch006

[13] Ganguly A, Das R, Setua SK (2020) Histopathological Image and Lymphoma Image Classification using customized Deep Learning models and different optimization

algorithms. 2020 11th Int Conf Comput Commun Netw Technol ICCCNT 2020. https://doi.org/10.1109/ICCCNT49239.2020.9225616

[14] Das R, Arshad M, Manjhi PK, Thepade SD (2020) Improved Feature Generalization in Smaller Datasets with Early Feature Fusion of Handcrafted and Automated Features for Content Based Image Classification. 2020 11th Int Conf Comput Commun Netw Technol ICCCNT 2020 1–5. https://doi.org/10.1109/ICCCNT49239.2020.9225439

[15] Saha M, Chakraborty C, Racoceanu D (2018) Efficient deep learning model for mitosis detection using breast histopathology images. Comput Med Imaging Graph 64:29–40. https://doi.org/10.1016/j.compmedimag.2017.12.001