Deep Learning for Digital Pathology ….. Using Representation Learning

Authors: Arockia Liborious, Gargi Biswas, Anish Anurag, Rik Das, Govind Kumar Jha, Sudeep Thepade

Next

Introduction/Background

Image processing and machine learning algorithms can be used to recognize tissue sample images

Why Digital Pathology via Image Analysis?

- ✓ Glass slide specimen analysis / storage
- ✓ Early detection can help save lives
- ✓ Allows multiple experts to analyze at the same time

Challenges:

- 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

Drawbacks:

Instrumentation & storage capacity of slides

Accuracy of estimates differ according to the pathologist’s experience

Next

Problem Statement

Computer vision techniques could be used to analyze tissue sample images, but the drawback is better generalization and faster applicability in real life with available data

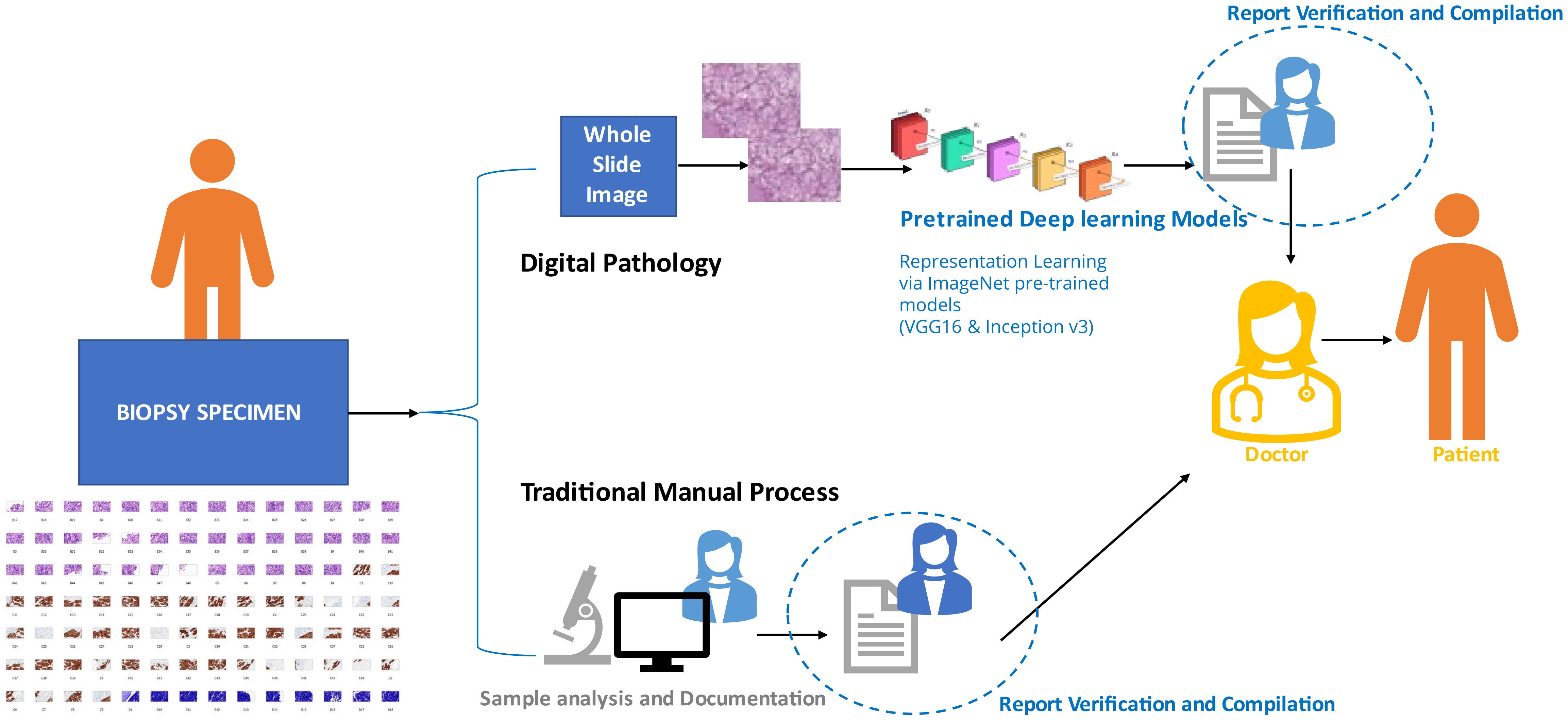
Aim

Propose a low cost, fast and improved tissue sample analysis technique to detect abnormalities in cell structure/growth using representation learning techniques

Objective:

- Modelling & Representation Learning
    - Use VGG16 & Inception v3 to extract feature maps and run SVM and Neural Network classifier
- Comparison Study
    - Experiment auto extracted features with SVM and Neural Network classifier to compare the

:: Proposed Methodology



Next

Researchers conclude that by correlating histological trends with nuclei study, protein, and gene expression, conducting exploratory histopathology image analysis, and performing computer assisted diagnostics (CADx), pathologists will be able to make better decisions

Literature Survey

- Researchers proposed the method of picture foreground extraction as an image segmentation to extract nuclei details of a cell that carries vital information in pathology. Overall, the process achieved an accuracy of more than 86%
- Whole Slide Image (WSI) technology has been steadily establishing laboratory standards as a method of digitizing pathology slides for more effective diagnostic, educational, and research purposes Researches show
- Researches show Deep Neural Networks considerably improve feature extraction process. Also handcrafted features help in extracting morphological and textural features

:: Analysis

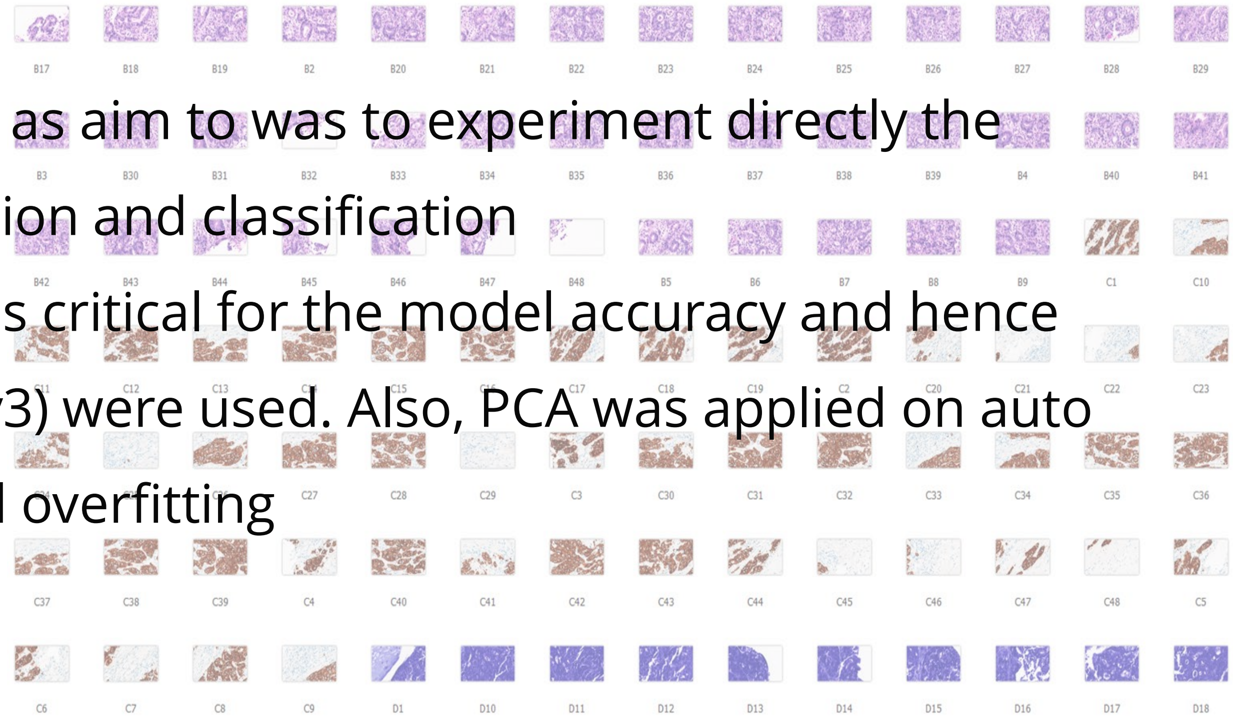
Next

Dataset: KIMIA Path960 which contain 960(=20×48) images of manually selected 48 regions of interest

EDA: Color / tissue staining information hold details of cell boundaries and extent of growth etc. and hence using an ImageNet pre-trained model can be helpful for high- & low-level auto feature extraction

No pre-processing technique was attempted as aim to was to experiment directly the handpicked images with auto feature extraction and classification

Feature Extraction: Right feature selection is critical for the model accuracy and hence two pretrained models (VGG16 & Inception-v3) were used. Also, PCA was applied on auto features to reduce dimension and help avoid overfitting



: : Results

Next

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

Note: All results were computed using 10-fold cross validation

Experiment	Features	Classifier	Accuracy	F1 Score	Number of Features	Run Time (mins)
1	VGG16	Neural Network	0.954	.954	4096	22.7
	VGG16	SVM	0.953	.953	4096	22.3
2	Inception v3	Neural Network	0.955	.955	2048	4.4
	Inception v3	SVM	0.947	.947	2048	3.5

Experiment	Features	Classifier	Accuracy	F1 Score	Number of Features	Run Time (mins)
3	PCA (VGG16)	Neural Network	0.863	0.856	150	18.5
	PCA (VGG16)	SVM	0.903	.904	150	18.3
4	PCA (Inception v3)	Neural Network	0.906	0.906	150	2.5
	PCA (Inception v3)	SVM	0.948	.948	150	2

Conclusion & Future Scope

Next

- 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 (> 85% in all experiments)
- PCs extracted from Inception-v3 auto features run ~87% faster than PCs of VGG16
- ANNs can overfit if training samples are less - a problem that SVMs do not have. Hence SVM seem to perform better and better generalizes the feature vector and thus 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