

Project Module 3 Paper
BMED/ ECE 4783

Submitted on May 2nd, 2016

Alex Coyle
Miguel Flores
Adam Kahin
Yuanda Zhu

Faculty Advisor: May D. Wang

I. Introduction and Problem Statement

Today, despite the advancements of medical science, cancer diagnosis is still a very imprecise process that often involves multiple physicians with differing opinions, and yet there are many cases of misdiagnosis. For these reasons, the development of effective diagnostic technologies is a crucial part of improving cancer care, and increasing survival rates world wide.

The aim of this project was to use image segmentation, feature extraction, and finally classification with the intent of creating an effective, automatic, and adaptable cancer diagnostic tool using image processing methods.

The images used in this project are kidney cancer tissues taken using microscopy. The photographed tissues are colored using H&E staining. This staining method detects and stains two primary structure types. Eosinophilic structures, which contain proteins, such as membranes and cytoplasm, stained pink. Basophilic structures which contain nucleotides stained blue/purple. Empty spaces remained white. All together, these features offer significant insight into the type and quality of tissue contained in an image.

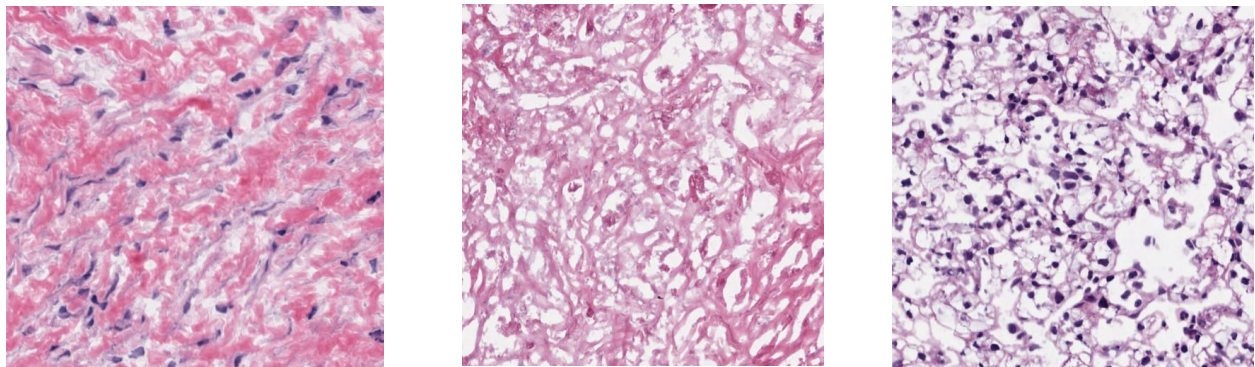


Figure 1. (a) Stroma. (b) Necrosis. (c) Tumor.

Stroma (left) is healthy tissue with appropriate quantities of pink and purple. This is because most of the cells in healthy tissue are in interphase, the period between cell division. Necrosis cells are dead or dying cells and are no longer undergoing the cell cycle, nor are they using basophilic structures like nuclei, RNA, or ribosomes to produce proteins or conduct any cellular functions. Therefore, necrosis images contain almost exclusively pink coloration. Tumor cells have uncontrolled proliferation, and are constantly dividing, so they are full of active nucleic acid structures. This is because

during cell division, chromosomes must be replicated. This makes it so that tumor cell images contain abnormal quantities of blue/purple compared to the other tissue types.

II. Module 1 - Image Segmentation

Literature Review for Module 1

Topic	Citation	Summary (5Ws, H)	Strengths and Weaknesses
Automatic nuclei segmentation	Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: A quantitative evaluation	S. Di Cataldo et al from Italy in 2009 proposed the automated segmentation through morphology-based approach. They compared their approach with active contours. Their motivation is to overcome the limitations of the existing methods with a fully automated morphology-based technique, and to provide a better segmentation accuracy.	The paper provides the idea of using nuclei segmentation for further feature extraction, but the iteration methods are too time costly.
Automatic nuclei segmentation	Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method	Aymen Mouelhi et al from Tunisia in 2012 presented a new automatic scheme to perform both classification of breast stained nuclei and segmentation of touching nuclei in order to get the total number of cancer nuclei in each class.	The paper provides the scheme of supervised method, but the given methods are too complicated.
Survey of Segmentation Methods	Image segmentation evaluation: A survey of unsupervised methods	Hui Zhang (2008) et al compiled and critiqued a large variety of segmentation evaluation algorithms. Using this paper, I gained the technical knowledge to further investigate evaluation methods and select meaningful performance metrics.	This paper was very broad, but very thorough. I served a good reference.
Theory of Segmentation	Toward Objective Evaluation of	This paper by Ranjith Unnikrishnan (2007) investigated a performance	This paper presented experimental methods and

n Evaluation	Image Segmentation Algorithms	metric known as Normalized Probabilistic Rand (NPR) index.	results that helped understand what a good evaluation metric should do.
theory of segmentation evaluation	An evaluation metric for image segmentation of multiple objects	Polak and Zhang (2009) investigated a sophisticated evaluation metric that accounted for several objects, not just one. This method is sensitive to both under segmentation and over segmentation.	This would be a great evaluation metric for our needs, but is too complicated to implement in our time-frame.
Supervised Learning Based Image Segmentation	Two learning based algorithms for cell image segmentation..	Presented a two step algorithm to first segment the cell nuclei from the back and then to use further processing that separates overlapping or touching nuclei.	Produced .4% and .2% of under segmentation and over segmentation
Supervised pixel classification method	Uses a strategy based on four steps in different color spaces.	Proposed an automatic segmentation scheme based on noise reduction, pixel classification in different color spaces, marker extraction, and color watershed growing to segment objects correctly..	Showed the importance of the choice of supervised classifiers and of color spaces
Automatic color segmentation system	Uses knowledge from pre-segmented reference images to normalize and classify pixels.	Provides an automatic color segmentation of histopathological specimens that are resistant to batch effects by incorporating knowledge from reference images.	Performs at 85% accuracy for systems that use color map normalization and reclassification
Automatic nuclei segmentation in breast cancer images	Developed an automated nuclei segmentation method in four steps.	Presented marker controlled watershed based technique for segmentation of cancer nuclei in combination with different processing steps	Detected and segmented tissue very well
Comparison of normalization algorithms for histopathological images	Five color normalization algorithms were compared using 204 images.	Found that the stain color normalization method using k-means clustering was found to be the best choice due to its high stain segmentation accuracy and lack of computational complexity.	Provided detailed results displaying difference between popular image segmentation techniques.

Detection of Tumor	Computer-Aided Detection of Kidney Tumor on Abdominal Computed Tomography Scans	D.-Y. Kim and J.-W. Park from Korea in 2004 implement a computer-aided detection system for kidney segmentation and kidney tumor detection on abdominal computed tomography (CT) scans. A gray-level threshold method was used to segment the kidney.	The students adopted the same idea of using histogram and gray-level thresholding as the author on supervised method.
Automated Cell Counting and Cluster Segmentation	Automated Cell Counting and Cluster Segmentation Using Concavity Detection and Ellipse Fitting Techniques	Sonal Kothari et al from Georgia Tech in 2009 presented a novel, fast and semi-automatic method for accurate cell cluster segmentation and cell counting of digital tissue image samples. Concavity detection at the edge of a cluster to find the points of overlap between two nuclei.	Would be beneficial to post-segmentation process.
Automatic color segmentation	Automatic Batch-Invariant Color Segmentation of Histological Cancer Images	Sonal Kothari et al from Georgia Tech in 2011 proposed an automatic color segmentation system which can normalize images to reduce sensitivity to batch effects. Image normalization significantly improve the accuracy of color segmentation.	Provided a detailed method to implement image normalization.
Threshold evaluation and selection methods	A threshold Selection Method from Gray-Level Histograms	Noboyuki Otsu in 1979 explored ways to assess the quality of and success of gray-scale images segmented with certain thresholds, such as intra-regional and inter-regional statistics.	Gave deeper understanding of intra and inter statistics such as eta.
color segmentation metric		In 1998, Borsotti investigated ways to assess the performance of color metric segmentation automatically without user feedback, seeking to make an objective measure that lines up with qualitative features of good segmentation	Sought to improve quantitative metrics
Improvements to a	A Color Texture Image	Guoying Liu, Pengwei li, and Yun Zhang from china added a Region-level	Gives a form of post and pre processing for

C-means clustering algorithm	Segmentation Method Based on Fuzzy c-Means Clustering and Region-level Markov Random Field Model	Markov Random Field Model to overcome the weaknesses of the fuzzy local c-means clustering algorithm in 2015	improvement upon a clustering algorithm
Segmentation by fuzzy inference in optimized color space	Color image segmentation using neuro-fuzzy system in a novel optimized color space	Somayeh Mousavi, Fazlolah Soleymani, and Navid Razmjooy in London during 2012 used a neuro-adaptive learning technique based on a sample image to improve segmentation or object detection	Gave insight into a supervised method of image segmentation
Histogram-based image segmentation	Human perception-based image segmentation using optimising of colour quantisation	Sung In Cho, Suk-Ju Kang, and Young Hwan Kim et al POSTECH in Korea presented an advanced histogram-based image segmentation method to enhance the quality of image segmentation during 2013	Provided information on using histograms within image segmentation
Use of image processing to quantify the amount of intramuscular fat in a sample	Predictive efficiency of distinct color image segmentation methods for measuring intramuscular fat in beef	A team in Santa Maria CA during 2015 presented a methodology they used to quantify the amount of intramuscular fat in beef through the use of image based analysis	Gave information on evaluation methods to determine different types of tissue
Use of adaptive filtering to reduce performance degradation of existing methods	Segmentation with saliency map using colour and depth images	Ji-Eun Lee and Rae-Hong Park at Sogang University in Seoul Korea during 2014 proposed a segmentation method using colour and depth images from which a saliency map would be generated and utilized adaptive thresholding to reduce performance degradation	Showed a form of adaptive thresholding and guided filtering to improve image segmentation

The aim of this module was to use supervised and unsupervised image segmentation algorithms in order to delineate important structures in kidney cancer cell images. Performing image segmentation allowed us to better see and understand the features within each tissue type by visually distinguishing the nuclei, cytoplasm, and background of the images. To perform image segmentation we used both a supervised and unsupervised method. Our unsupervised method was done through K-means Clustering while our supervised method was done through Histogram Thresholding.

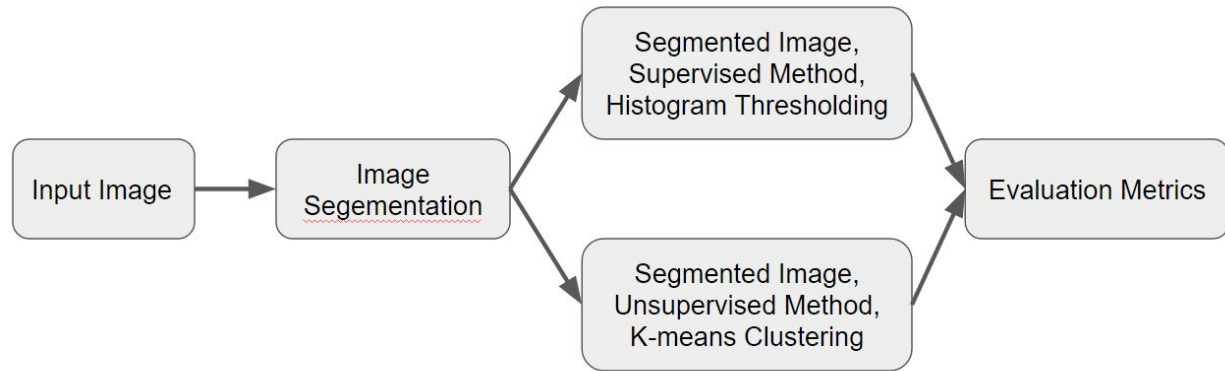


Figure 2. System Flow of Image Segmentation.

Our performance metrics showed that our Unsupervised method was a better method of image segmentation than our supervised method. This was a surprising result, but is likely due to poor choice of reference nodes in our supervised method.

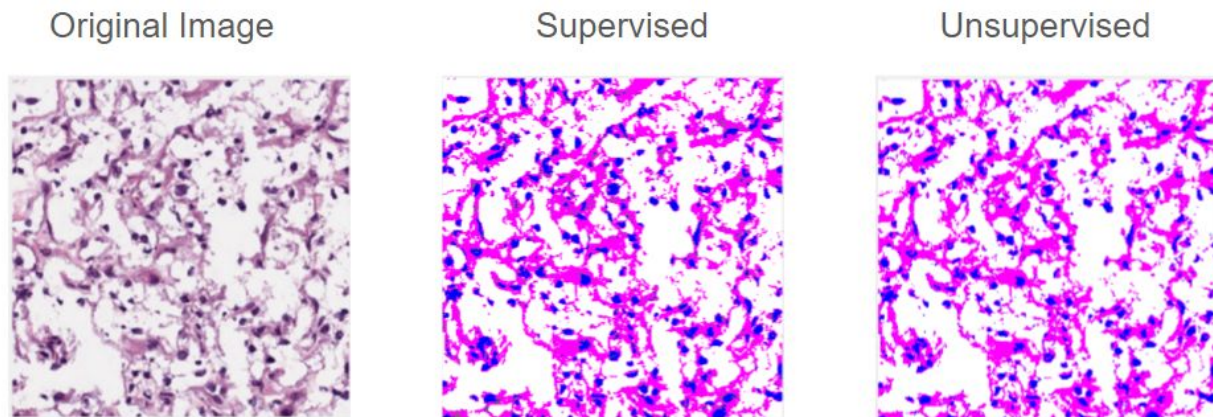


Figure 3. Example of Segmentation from Each Method.

III. Module 2 - Feature Extraction

Literature Review for Module 2

Author(s), Year	Title	Summary
Loris Nanni, Sheryl Brahnam, Alessandra Lumini, 2010	A local approach based on a Local Binary Patterns variant texture descriptor for classifying pain states	The authors compared several texture descriptors based on local binary pattern (LBP) and proposed the new texture descriptor: the Elongated Ternary Pattern (ELTP) and the Elongated Binary Pattern (ELBP). For their experiment, ELTP was proved to generate the best results.
Laura E. Boucheron ,2008	Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology Images for Detection and Characterization of Cancer	Provides a detailed description of techniques for Higher Level Image analysis. It details many concepts of object level features for image classification.
Włodzisław Duch, 2004	Comparison of Feature Ranking methods based on Entropy	Gave an overview of entropy based ranking metrics, and gave a theoretical framework for feature ranking.
R.Venkata Ramana Chary, Dr. D. Rajya Lakshmi and Dr. K.V.N Sunitha, 2012	Feature Extraction Methods for Color Image Similarity	Provides insight to a methodology of grouping with threshold values and sub grouping through the use of RGB color combinations

M. Mudrov´a, A. Proch´azka,	Gave us an understanding of the use and application of Principal Component Analysis	Gave us an understanding of the use and application of Principal Component Analysis
-----------------------------	---	---

The aim of this module was to perform feature extraction through three different primary methods: color, texture and morphology. This gave us insight into the distinguishing characteristics of each tissue type which allowed us to diagnose new images by comparing parameters we find to have the most significance in determining the different types of tissue. These parameters were first extracted from each image and then ranked and reduced. The top ten parameters from each method can be seen in Table 1 the methods used to extract these features is shown in Figure 4.

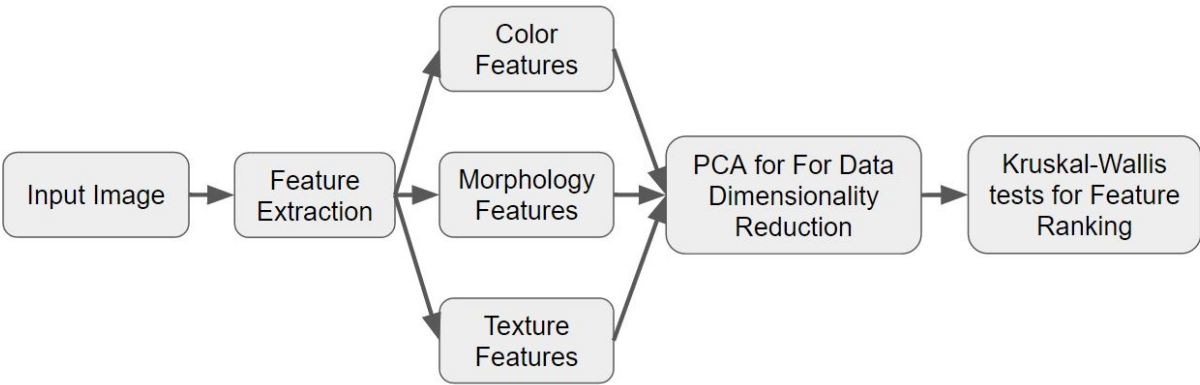


Figure 4. System Flow of Feature Extraction.

Table 1. Top 10 features from each method.

Ranking	Color	Texture	Morphology
1	Full image, red channel mean	Energy 90 deg	Standard Dev. of Area
2	Full image, red channel kurtosis	Homogeneity 315 deg	Eccentricity

3	Full image, red channel standard deviation,	Contrast overall	Std. Dev of Major Axis Length
4	Full image, red channel variance	Contrast 225 deg	Std. Dev of Minor Axis Length
5	Nucleus, red channel skew	Contrast 45 deg	# of Nuclei
6	Full image, red-green channel standard deviation	Contrast 315 deg	Mean Area
7	Full image, red-green-blue channel skew	Contrast 0 deg	Mean of Minor Axis Length
8	Full image, red channel skew	Contrast 180 deg	Radius
9	Nucleus, red channel, mean	Contrast 270 deg	Mean of Orientation
10	Nucleus, red channel kurtosis	Contrast 90 deg	Mean of Major Axis Length

IV. Module 3 Literature review

Topic	Title	Summary (5Ws, H)	Strengths and Weaknesses
FDA	Fisher Discriminant Analysis with kernels	Explains non-linear classification techniques based on Fisher's discriminant using kernels allow efficient rotation of the feature space.	Although this method is one of the standard linear technique for data analysis. However, linear methods are often too limited.
SVM	Introduction to SVM	Describes the basic functions in Matlab that implement support vector machines and how to use them.	Great resource for implement simple classification, but lacks details on how to use more advanced parameters for the classification model..

SVM	Cancer Classification SVM and RVM based on is of Variance Features	Develops a method to find the smallest set of genes that can ensure highly accurate classification	There SVM method performed well, and provided a great flow graph diagram.
NN	Combining the Results of Several Neural Network Classifiers	Galina Rogova 1994 This paper presents a combination method based on the Dempster-Shafer theory of evidence, which uses statistical information about the relative classification strengths of several classifiers.	The method adopted allows 15-30% reduction of misclassification error compared to the best individual classifier.
NN	Artificial Neural Networks: A Tutorial	Anil K. Jain from Michigan State University, Jianchang Mao and K.M. Mohiuddin from IBM Almaden Research Center 1996 This tutorial discusses the motivations behind the development of ANNs, describes the basic biological neuron and computational model, outline network architectures and learning process, and presents most commonly used ANN models.	This paper is easily accessible to the readers who have no previous knowledge about artificial neural networks. The paper concluded with pattern recognition, a successful ANN application.
NN	A Methodology to Explain Neural Network Classification	Raphael Feraud, Fabrice Clerot, 2011 The paper explained neural network method by a multi-layer perceptron. The authors defined a clustering of data built from hidden layer representation after the model was trained with relevant variables.	The saliency measurement allows to select the smallest set of input variables to simplify a neural network, when the input variables are uncorrelated.
KNN	Distance Metric Learning for Large Margin Nearest Neighbor Classification	Kilian Q. Weinberger, John Blitzer, and Lawrence K. Saul The metric is trained with the goal that the k-nearest neighbors always belong to the same class while examples from different classes are separated by a large	One of the oldest and simplest methods for pattern classification. Yields competitive results, but its performance depends crucially on the distance metric used to identify

		margin.	nearest neighbors.
KNN	Using the K-Nearest Neighbor Algorithm for the Classification of Lymph Node Metastasis in Gastric Cancer	Chao Li, Shuheng Zhang, Huan Zhang, Lifang Pang, Kinman Lam, Chun Hui, and Su Zhang in 2012 China Used the K-nearest neighbor classifier along with machine learning to distinguish lymph node metastasis from non-lymph node metastasis	Determined k-nearest neighbor on GSI and CT scans was an effective method for classification with an accuracy of 96.33%. Required a lot of pre-processing for such a level of accuracy
Decision Tree	Comprehensive Decision Tree Models in Bioinformatics	Gregor Stiglic, Simon Kocbek, Igor Pernek, Peter Kokol 2012 Used machine learning techniques to improve the classification and tuning of a decision tree classifier. Study demonstrated suitability of visually tuned decision trees.	Demonstrated that using predefined boundaries gave very good classification performance. Used primarily for binary class attributes
Decision Tree	A Survey of Decision tree Classifier Methodology	S. Rasoul Safavian and David Landgrebe 1991 This survey presents current methods for DTC designs and the various existing issues. After considering potential advantages of DTC's over single-state classifiers, the subjects of tree structure design, feature selection at each internal node, and decision and search strategies are discussed. Some remarks concerning the relation between decision trees and neural networks (NN) are also made.	This survey has brought the disparate issues in decision tree classifiers closer together and perhaps motivate some new ideas, provided a more unified view of decision tree classifiers and caution the "casual" users of these methods of the possible "pitfalls" of each method.

V. Module 3 - Classification

The goal of this module was to use three different classifiers to give an accurate diagnosis of tissue types. This was accomplished through the use of K-nearest

neighbor, support vector machine, and neural network algorithms, in addition to the use of k-fold and leave one out cross validation.

Each of these algorithms were given the top 10 features from each method from our feature extraction module. The rationale behind this choice of using only 30 features was due to the diminishing returns we were seeing in increased accuracy compared to the increasing in computational times. Using a 30 feature total provided sufficient accuracy with a low impact computational time.

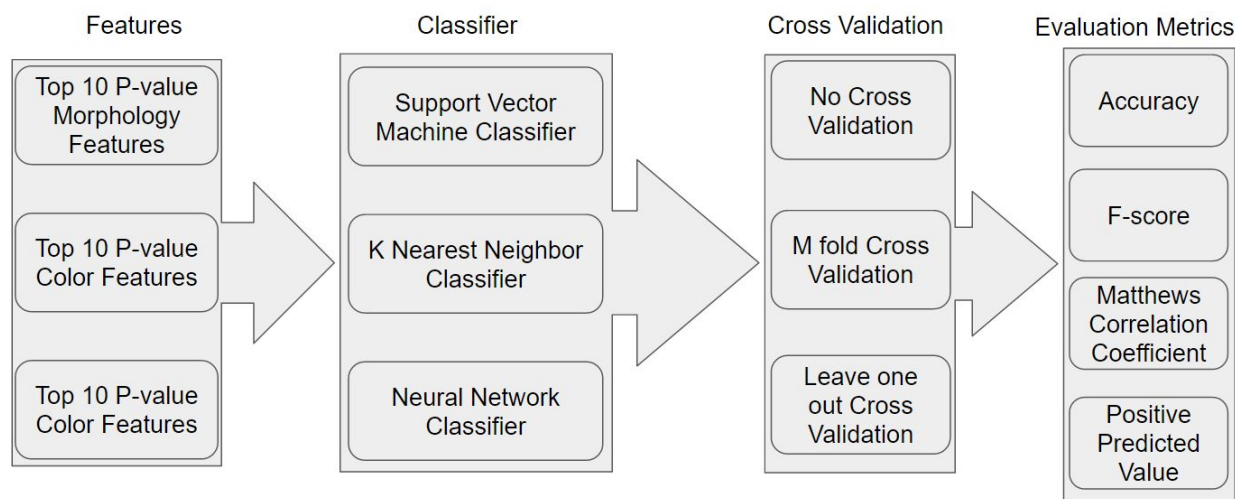


Figure 5.. System Flow of Classification.

1. KNN

Our first method of classification was K-nearest neighbor. This method compares the data point of an image in the testing set to same kind of data point from every image in the training set. It then finds the value in the training set that is closest to the testing value and gives us the location of this closest value. Based on this location we assign a value of 1, 2, or 3 as a prediction as seen in Table x, corresponding to Stroma, Necrosis, and Tumor respectively.

This methodology is then repeated a total of 30 times per image, once for each feature based on our top 30 features from module 2, giving a tissue prediction per feature. The mode of these 30 predictions is then taken as the final prediction for the classifier. The KNN method is then done twice more, but cross validated through K-folding and the leave-one-out method.

Table 2. Sample output of the KNN Method. Red is an incorrect prediction

Img #	Stroma	Necrosis	Tumor

61	1	1	3
62	1	2	3
63	1	1	3
64	1	2	3
65	1	3	2
66	1	2	3
67	1	2	3
-	-	-	-
100	3	2	2

2. Neural Network

Our second method of classification was Neural Network. Matlab built-in GUI “nnstart”> “Pattern Recognition” has been used to help us select data, create and train a network, and evaluate its performance using cross-entropy and confusion matrices.

A 20*300 matrix was adopted as the input. 30 selected features of 300 images were included in this matrix. The target output is a 3*300 matrix that contains only 1s and 0s. In each column, one 1 and two 0s exist. The binary number 1 matches the specified image with the corresponding image type.

Table 3. Target output matrix. Only the 1s and 0s are in the actual target output matrix.

	N1	N2	...	N99	N100	S1	S2	...	S99	S100	T1	T2	...	T99	T100
N	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

Cross validation is approached through changing the percentage of validation and testing set. The training set would always be 60% of the entire image data sets. In original test, validation set has 10% samples and test set has 30%; for k-folding cross

validation, the validation set has increased to 20% and test set decreased to 20%; for leave-one-out cross validation, the validation set is 30% while test set 10%.

Due to the mechanics of Matlab GUI nnstart, different prediction results are generated even with exactly the same inputs and settings.

Table 4. Prediction output matrix. Only the 1s and 0s are in the actual output matrix.

Wrong Prediction Results

	N1	N2	...	N99	N100	S1	S2	...	S99	S100	T1	T2	...	T99	T100
N	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0
S	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0
T	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1

3. Support Vector Machine

The third method chosen for classification was Support Vector Machine. The classification model was built using a function in Matlab called 'fitsvm()'. The input to this built-in function was a training set, which comprised the top ranking features found in module 2, and a labeling set, which comprised of one of two labels distinguishing two groups in the set; those that belong to a specific class of tissue and everything else. Since the function was only able to classify/predict two groups at a time, three models were created for each specific class (Stroma, Necrosis, and Tumor). Once each model was created, the model as well as a testing set was input to a SVM prediction function called 'predict()'. The output of the prediction function was an array of prediction results as well as a corresponding score from 0 to 1. The prediction chosen for a particular image in the testing set was the prediction with the greatest score. A flow diagram is shown in Figure 6.

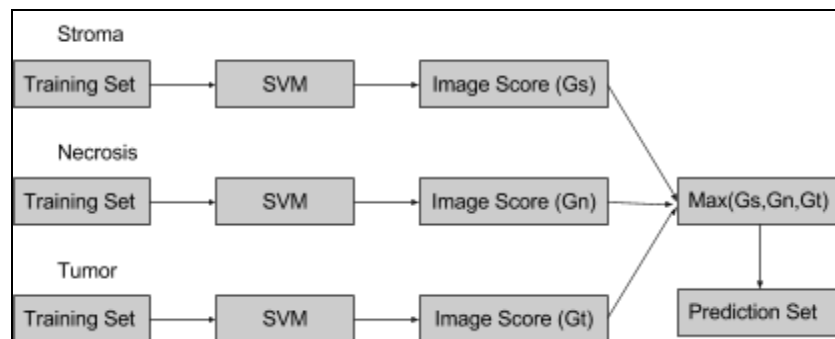


Figure 6. Flow diagram of SVM classification

4. Cross Correlation

Our cross validation was accomplished through both the k-folding method and the leave one out method. In our initial test, also known as the holdout method, we used images 1-60 from each class as our training set and images 61-100 as our testing set.

For our k-folding method, the testing set was divided into 4 sets of 10. Each set of ten is then tested against our initial training set and the other 30 used as our validation set. This test is then repeated a total of 4 times in order to test every image in our testing set.

The leave-one-out method is essentially our k-folding method, but taken to the absolute extreme. Using the same initial training set, the k-fold method is turned into the leave-one-method by using 1 image as the testing set and the remaining 39 as the validation set. This method is then repeated a total of 40 times to test every image in our testing set.

5. Performance Metrics - Results

When evaluating the performance of our classification methods, more than one metric is required for a comprehensive understanding of our method's performance. Four different performance metrics were chosen. Accuracy, F-score, Matthews Correlation Coefficient (MCC), and Positive Predictive Value (PPV).

Accuracy is a very easy and intuitive measure of the quality of a classification method. Accuracy is the proportion of correct true positives over total size of the set, N. This gives a number between 0 and 1.

$$Accuracy = \frac{TP}{N}$$

F-score is a method of measuring the efficacy of binary classification. It assigns a score from 0 to 1 that characterizes the accuracy of a classification method by providing the harmonic mean of precision and recall. In this case, precision is the fraction of retrieved instances that are **true**, while recall is the fraction of **true** instances that are retrieved. The F-score can be calculated from the confusion matrix using the equation below.

$$F = \frac{2TP}{2TP + FP + FN}$$

Matthew's correlation coefficient is a measure of tests predictive value. It designates a value from -1 to +1, indicating the quality of a test. A score of 0 indicates a worthless test. A score of 1 indicates a perfectly accurate test, while a score of -1 indicates a perfectly inaccurate test. The equation below can be used to calculate the MCC from the confusion matrix.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(FN + TP)(TN + FP)(TN + FN)}}$$

Positive Predictive value is the proportion of true positives to the total number of positive predictions. This metric quantifies the value of a test's ability to correctly detect the disease state.

$$PPV = \frac{TP}{TP + FP}$$

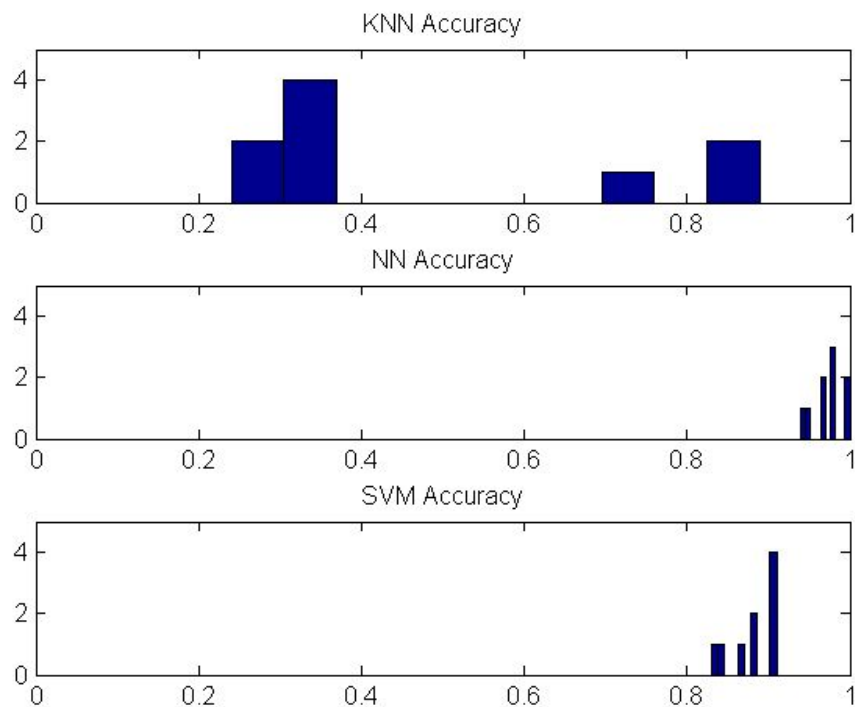


Figure 7. Graph of Accuracies

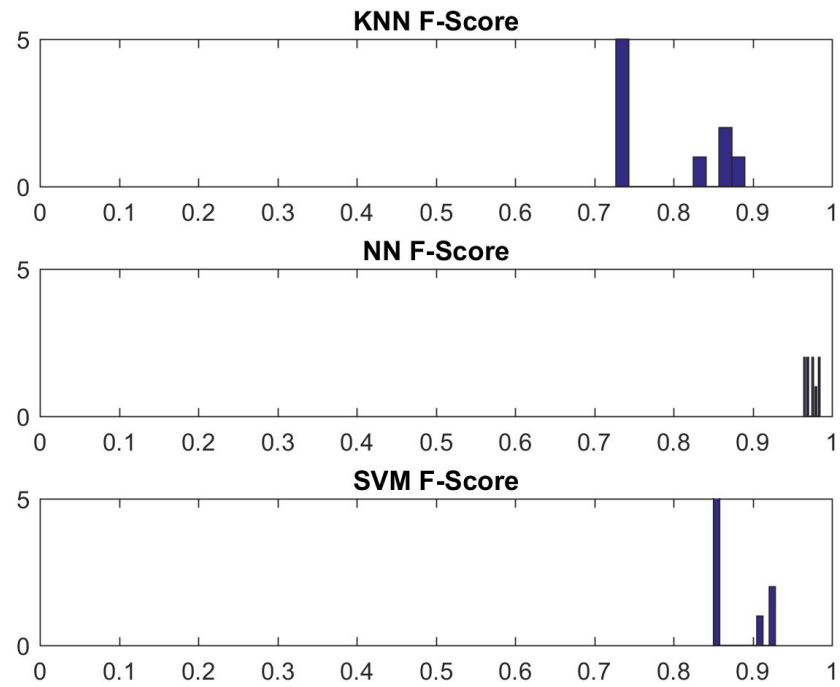


Figure 8. Graph of F-Scores

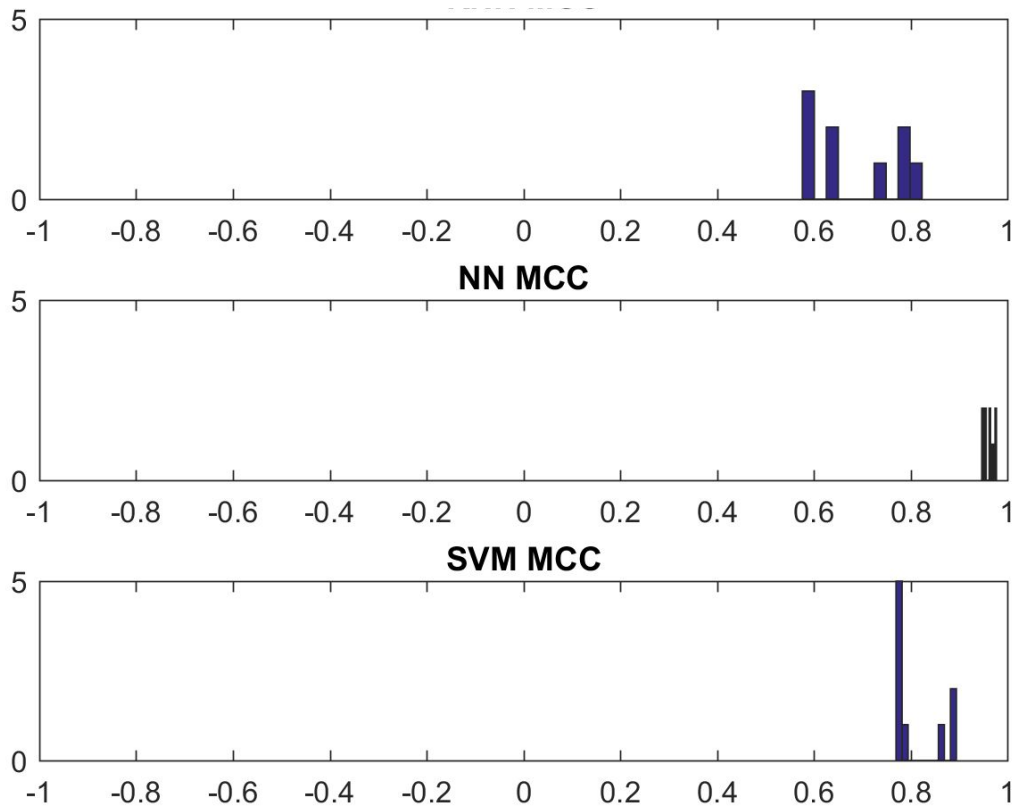


Figure 9. Graph of Matthew's correlation coefficient

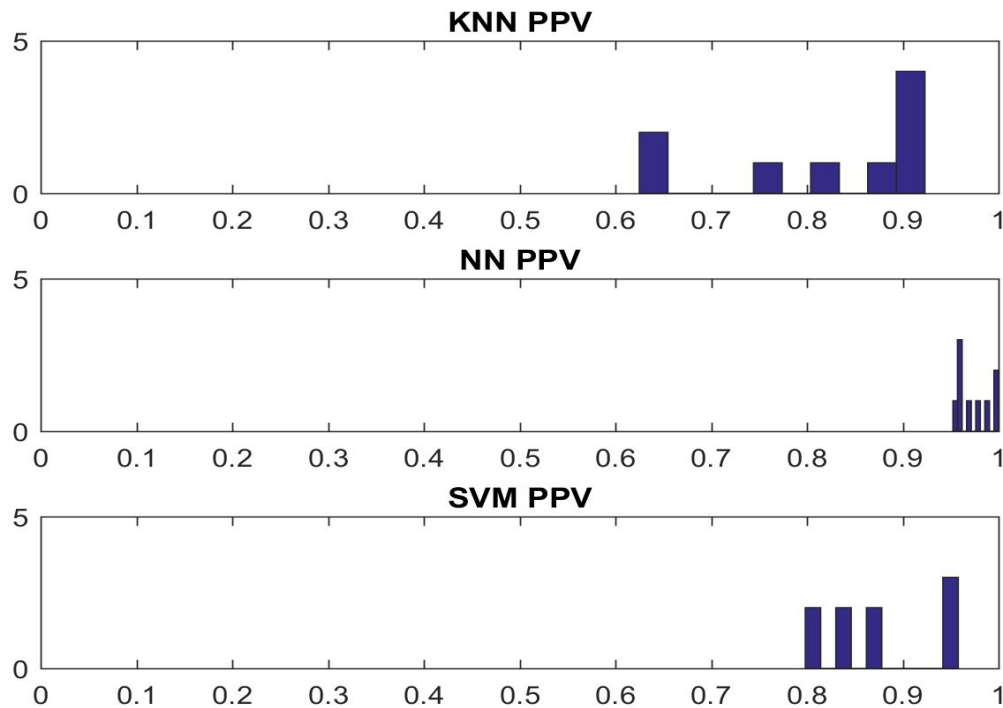


Figure 10. Graph of Positive Predictive value

VI. Conclusions

The majority of the KNN method inaccuracies came from the mis-prediction of Stromas as Necrosis and Necrosis as Stromas. Our KNN method did, however, excel in its ability to accurately distinguish tumors from the other two tissue types with an accuracy of about 91%. Using cross validation in this classifier had a significant impact on its overall accuracy.

The accuracy of our neural network method is nearly equivalent in its ability to predict the three classes of tissue image. This method's overall accuracy fell in the range from 90% to 97%. Cross validation did improve this method's accuracy, but not by a large amount.

The SVM method was able to distinguish between each class of tissue with an accuracy around 88%. When cross validation was ran on this method, it showed a minimal in accuracy. Based on this cross validation is not necessary past a k-folding method as it only increases the computational time.

Overall, neural network gave the best prediction result out of our three methods, in addition to being the most accurate for each individual class.

VII. Future Work

A few steps have been laid out for future work. The identification of a greater number of significant features for each classifier is the first major step in this process. The addition of more features will lead to better classification results by giving us more potential points for comparison. In conjunction, each feature shall be weighted in our classifiers based on its P-value rank. Weighting our features would cause the most significantly different features to have the greatest impact on our predictive models. Currently we treat all of the features equally during our classification which could slightly skew the results. The last step is to fine tune variable parameters in each of our classifiers to achieve a better overall performance.