Computer-Aided Diagnosis of Lesion in Surface Tissue Photographs using Image Processing and Machine Learning

Suleiman Mustafa (Completed doctor's program in September 2019)

Main Objectives

1. Propose a novel computer-aided diagnosis system from macro-images
2. The system can assist medical personnel in identifying melanoma and cervix cancer
3. The system is based on image processing and machine learning
4. Aim to achieve higher accuracy of detection compared to humans visual inspection of cervix cancer 45–79% and
5. Existing methods achieved with small data for melanoma is around 73.4-93.6% and cervix cancer is around 65.1-82.5%.

Methodology

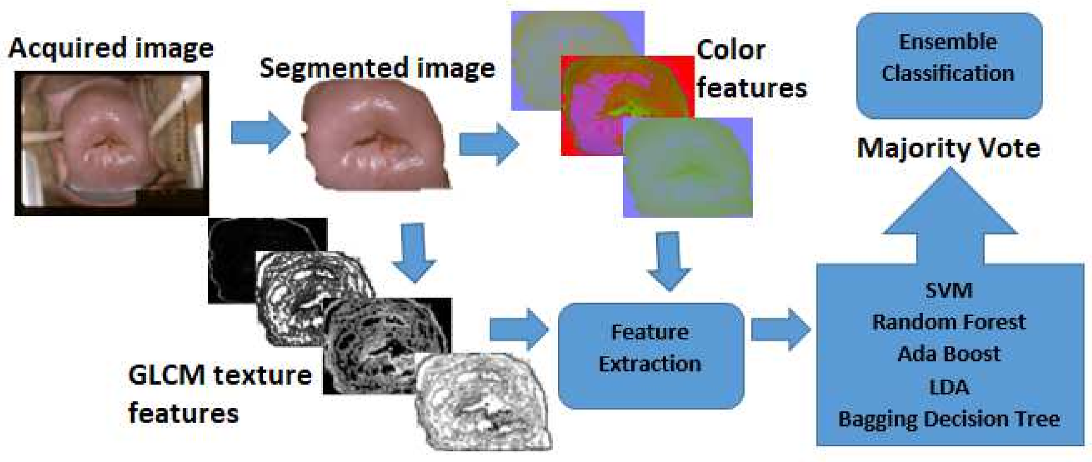

Figure 1: Approach for Cervix Cancer Classification

Approach for melanoma includes pre-processing to enhance the images for improving segmentation. Moreover, features based on shape, color and boundary are calculated.

Image Acquisition

Figure 2: Positive Melanoma samples

200 skin images acquired from on-line database sources: https://www.dermquest.com/ and http://www.dermis.net. ABCDEs rules is used to detect if mole is melanoma or not. ABCDE stands for Asymmetry – one half matches another, Border – uneven borders, Color – variety of colors, Diameter – grows above 6mm, and Evolving – change in size

Figure 3: Detecting the ABCD's of Melanoma

3,637 cervix images through Technology Transfer Agreement with Centre of the National Cancer Institute (NCI), USA. Images of normal (no risk) and CIN1 (low risk) treated as "Negatives", Images of CIN2, CIN3 and CIN4 (medium risk, high risk and cancer risk, respectively) as "Positives".

Segmentation

Using GrabCut method, in melanoma rectangle region set based on 75% output image dimension. Detected correct melanoma foreground pixels accurately about 91.3%.. In cervix cancer, GrabCut with mask to segment affected part without manual input. Detected correct cervix foreground pixels accurately about 87.1%. For the segmented melanoma images 15 lesions properties to follow the ABCDE'S rule are extracted. Moreover, in cervix cancer 46 features based on color and texture to follow VIA are extracted.

Feature Extraction and Selection

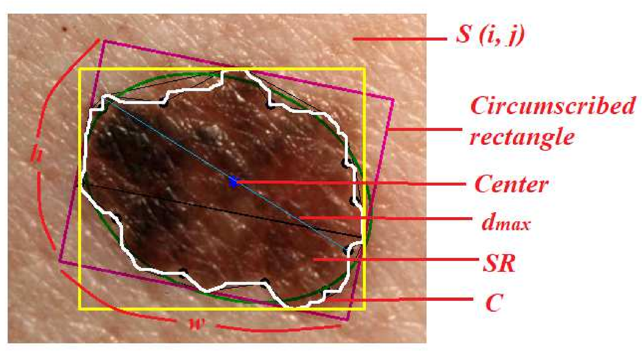

Figure 4: Segmented region from which features are calculated

Sequential backward selection (SBS) is applied to select a subset of the most useful features from the full set of 15 and 46 features in melanoma and cervix cancer respectively. Starting from the full set of features $X = \{x_1, x_2, \ldots, x_{15}$ and $x_{46}\}$ and returns a subset of features $X_{N_f} = \{x_j | j \in 1, 2, \ldots, 15$ and $46\}$; where $N_f (\leq 15$ and $46)$ is the pre-defined number of selected features. SBS determined only six of extracted melanoma features are enough to classify melanoma. Furthermore, SBS also determined only 10 - 13 of extracted cervix features are enough to classify cervix cancer.

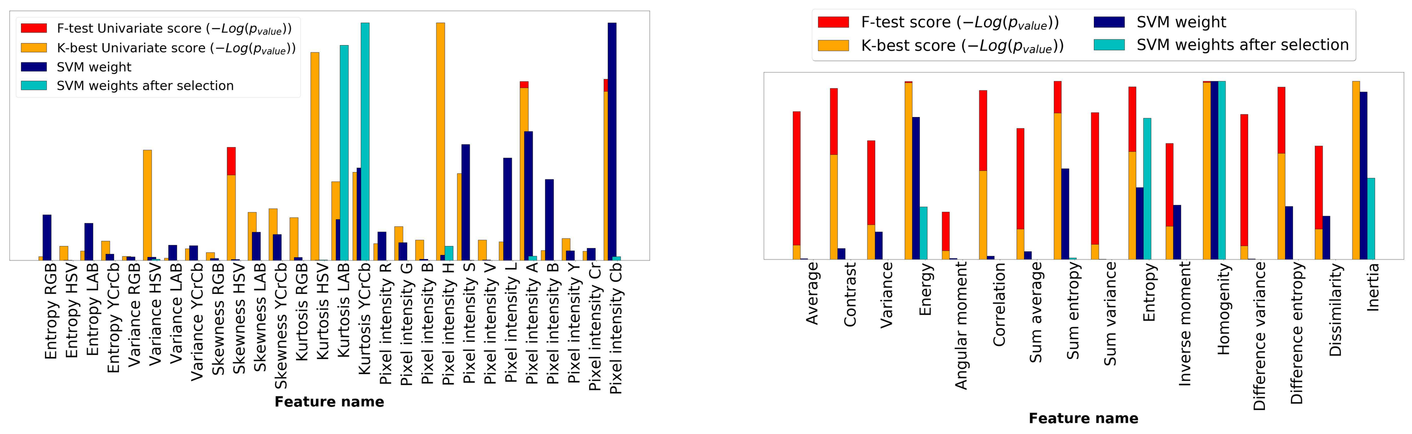

Figure 5: Univariate selection to rank the importance of cervix features

Classification

Support vector machine (SVM) generalizes well with unseen data and smaller data samples. Gaussian radial basis function (RBF) kernel for improved performance than linear or polynomial.

Ensemble method based on majority voting improved performance. Where the six-dimensional input feature vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_6]^{\mathrm{T}}$ extracted from an input affected skin image. The class of $\boldsymbol{x}$ is determined as "positive" when the condition

$$\sum_{t=1}^{T} d_{t,P}(\boldsymbol{x}) > \sum_{t=1}^{T} d_{t,N}(\boldsymbol{x}) \tag{1}$$

is satisfied, and otherwise, as "negative". Where $d_{t,P}$ is 1 or 0 depending on whether t^{th} classifier chooses positive, or not, respectively. $T = 3$ or 5 since we use 3 and then 5 machine learning algorithms. These are Support Vector Machine with Gaussian radial basis function (SVM RBF), AdaBoost (AdB), Random Forest (RF), Gradient Boosting (GB) and Bagging with Gaussian Naïve Bayes (BGN). 10-fold cross validation is used with images split to 70% and 30% i.e 140 training and 60 testing in melanoma and 2545 training and 1092 testing in cervix cancer.

Results for Melanoma

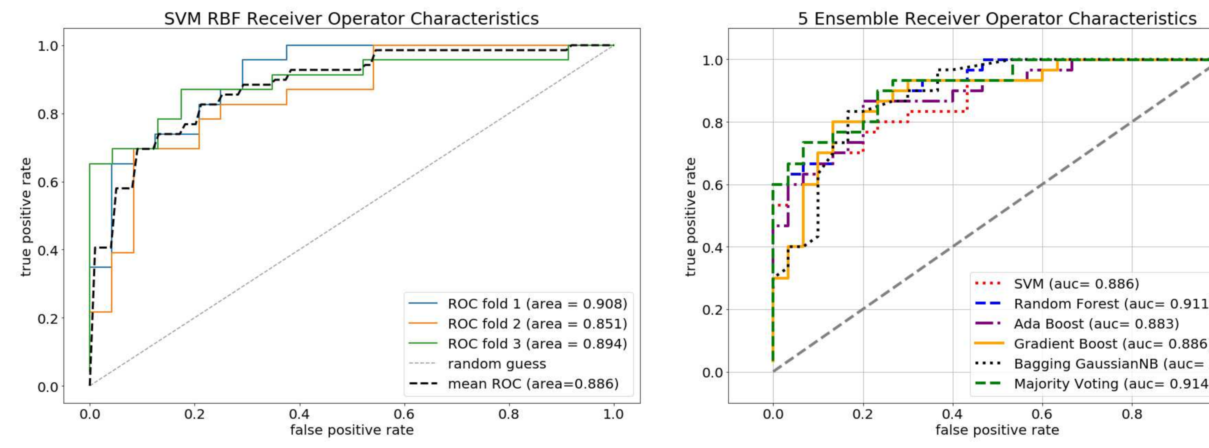

Figure 6: ROC for melanoma with SVM and Ensemble method

Classifier	Accuracy	F1	AUC
SVM RBF	86.6%	0.910	0.886
3 Ensemble	87.3%	0.877	0.900
5 Ensemble	85.7%	0.860	0.914

Table 1: Summary of melanoma results

Evaluation of ensemble method determined the best combination of 3 classifiers to be (SVM RBF, RF and AdB). Using Receiver operating characteristic (ROC) to show the diagnostic ability and AUC is the area under the curve. Area under the curve (AUC) above 0.90 achieved, scores improve from 0.886 with SVM, to 0.90 and 0.914 with 3 and 5 Ensemble respectively.

Results for Cervix Cancer

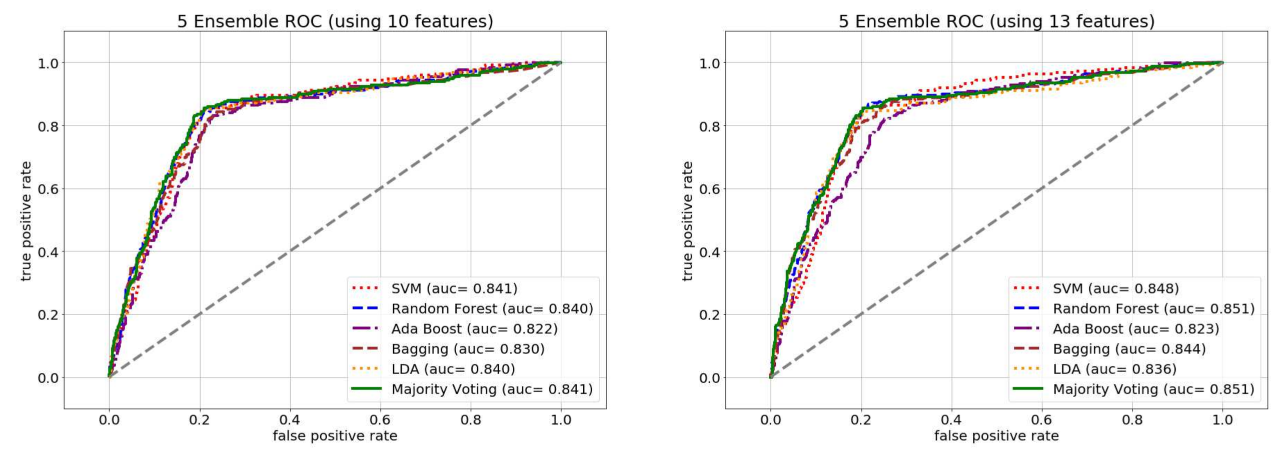

Figure 7: ROC for cervix cancer with 10 and 13 features

Classifier	Accuracy	F1	AUC
SVM RBF	83.6%	0.827	0.848
3 Ensemble	83.4%	0.838	0.852
5 Ensemble	83.9%	0.834	0.851

Table 2: Summary of cervix results

Selecting consistency-based and correlation-based features, 2 sets with 10 and 13 features are obtained. AUC scores of about of 0.84 obtained with 10 features. Slightly improved AUC above 0.85 achieved with 13 features. Results for 3 and 5 ensemble have no significant difference.

Conclusions

- Our segmentation method is effective to separate melanoma and cervix cancer lesions
- Few sets of 6 useful features in melanoma and 10-13 in cervix cancer are sufficient
- Achieved accuracy of 86.6% for melanoma with SVM RBF and accuracy of around 83.0 – 85.1% for cervix cancer
- Ensemble methods improve accuracy slightly: Combination of SVM-RBF, AdaBoost and Random Forest improved accuracy from 86.3% to 87.3% for melanoma Combination of SVM-RBF, AdaBoost and Random Forest improved accuracy from 85.1% to 85.2% for cervix cancer compared to 79% for VIA by doctor