CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

William C. Hang

Project Number

S1408

Project Title

Semantic Multilayer SVM: Novel Artificial Intelligence Applied to Prostate Cancer Grading and Breast Cancer Diagnosis

Abstract

Objectives/Goals

We develop a novel Support Vector Machine (SVM) learning algorithm called Semantic Multilayer Support Vector Machine (SMLSVM) that can take groups of low-level numerical data and translate them into higher-level concepts. We also develop an image analysis toolkit, Procist, to translate a prostate histology image into numerical data. We apply Procist and SMLSVM to prostate cancer grading based on two Gleason Grades, a novel development.

Methods/Materials

SMLSVM generates fewer, more valuable higher-level features out of many, lower-level numerical features. These capabilities allow it to reduce the curse of dimensionality and the semantic gap.

Procist is a computer vision algorithm that can generate numerical data that characterize the spatial arrangement and irregularity of cells, nuclei, stroma, lumens, and other tissue elements.

All algorithms were developed in MATLAB, and were tested on 20 prostate cancer histology images from the Johns Hopkins Medical Institute (JHMI) without patient info. We further validated SMLSVM alone on the University of Wisconsin Breast Cancer Dataset to differentiate between 699 malignant and benign tumor cases.

Results

We performed 10-fold cross validation on both datasets, and preliminary testing indicates that my algorithms outperformed current SVM approaches on both datasets. Results from further testing will be updated and presented during the fair.

Conclusions/Discussion

The objectives of this project have been achieved. SMLSVM and Procist could classify cancer malignancy and cancer severities with very high accuracies.

Few studies have been performed on Multilayer SVMs, and we present new research into SVM technology, artificial intelligence, and deep learning. Furthermore, by differentiating cancers based on two Gleason Grades, we present novel contributions to computational histopathology, where current approaches only differentiate between one Gleason Grade and are much less clinically accurate.

A provisional patent is pending on SMLSVM. I am currently working with JHMI to secure a much larger dataset for a better validation of my algorithms' capabilities.

Summary Statement

We present a novel artificial intelligence architecture that uses the collective power of SVMs to perform data abstraction, and apply this method along with newly developed image processing algorithms to autonomously grade prostate cancers.

Help Received

Dr. Robert Veltri from Johns Hopkins Medical Institude supplied necessary testing images and support, David West provided encouragement and advice.