DREXEL UNIVERSITY Computer Science College of Engineering

REThink CS @Drexel

Research Experiences for Teachers in Engineering and Computer Science:
Machine Learning To Enhance Human-Centered Computing

Research Proposal

Project Title: Development of documentation and training tools for the reproduction and dissemination of histological research results.

Name(s): Laurence Liss

Faculty Mentor(s): Dr. Mark Zarella and Dr. David Breen

Affiliates: *other persons (e.g., student assistants) involved in this research.*
Jason DeFuria

Abstract: *1-2 paragraphs explaining what your research is to accomplish, and how it relates to the mentor's research.*
In order to promote the continued adoption of published algorithms developed by Dr. Zarella I intend to build a documentation and tutorial repository of said algorithms. We hypothesize that complexity and lack of clear instruction may be hindering some users. The project aims to create a web-based instruction manual that can be continually updated and managed by Dr. Zarella, Dr. Breen, and those working in their labs. A key component of this work will be reproduction of the results along with detailed documentation. A further goal is to provide the algorithms in a language, such as Python. This secondary goal, to begin after successful reproduction and creation of the documentation website, will be to port the algorithms to Python and compare the results of both the original and new versions. I hope that distributing the algorithms in Python will help bring them to an audience that may not otherwise investigate these findings. Using standard web analytics and working with Drexel's web teams, it should be possible to understand the level of interest in different versions of the software which could be helpful in making future decisions for how development should proceed to ensure the largest possible audience.

Social Relevance and Potential Impact: *1 paragraph explaining how your research has potential social relevance.*
Providing research data and results in an easy to use manner is an often overlooked aspect of the research and publication process. Yet, access to data, technical requirements, and ways to reproduce the results are key to the furthering of scientific knowledge. Further, in the context of Dr. Zarella's work, the research can help clinicians perform diagnostic work.

Objectives/Timeline: *subtasks/milestones and expected dates for achieving them.*
Week 1: Project research and proposal.
Week 2: Reproduction of results along with detailed documentation.
Week 3: Create web-based instructions and code samples.
Week 4: Begin porting programs to Python.
Week 5: Continue Python development work.
Week 6: Finalize results. Document findings. Prepare poster session.

Resources Needed: *describe here any additional software, hardware, reference texts, scholarly works, etc., needed.*
MatLab license or access to computers with MatLab installed.

Programming Requirements: *describe here any new code that will need to be written to meet your research goals. This may be for system prototyping, data analysis, or other aspects of your project. Indicate whether you expect to need programming support from the project staff.*
Python versions of the original algorithms will be a goal of the project. Some assistance would be helpful as I have limited familiarity with Python-based machine learning and scientific computing tools.

New Knowledge Needed: *describe here any additional learning you need in order to meet your goals.*
In order to analyze Dr. Zarella's work and convert it to another programming language, I will have to learn some of the fundamental techniques of machine learning algorithms and how they apply to the work being formed by the algorithms.

New Classroom Materials: *describe here the instructional materials you expect to produce.*
Through this project, I will be gaining direct experience with machine learning and its application to real-world problems. This should provide good baseline to introduce these topics in the classroom. Beyond this, I think the most interesting fundamental concept that is essential to the understanding of Dr. Zarella's work is the idea that the data contained in his normalized image files is not data representing color but is in fact a representation of structural elements. This property allows the images to be recolored in whatever manner is helpful to the human viewer. But this concept could be applied to many different types of images or map-based data. I envision developing a lab that would allow students to build an image viewer based on this concept and allow them to choose different colors for certain structures.