**Cox-Lauf – MSDS692 Practicum I Project Proposal**

**1. Name, Contact info (e.g. email/phone):**

Elan (“Lonny”) Cox-Lauf, [ecoxlauf@regis.edu](mailto:ecoxlauf@regis.edu), [lonnyjcox@gmail.com](mailto:lonnyjcox@gmail.com), (303) 522-0748

**2. Title of the project:**

IntestinalParasitic EggClassification in Microscopic Images through Deep Learning

**3. High level description of the project: what question or problem are you addressing?**

In this project I will be addressing the feasibility of using Deep Learning techniques (likely a CNN model) to classify the type of parasitic egg using microscopic images from human samples. Intestinal parasitic infections (IPIs) in patients can be misdiagnosed, and this misdiagnosis can result in negative outcomes, such as the incorrect prescriptions for drugs (Bahrami et al., 2018).These IPIs are more prevalent in developing countries where there is limited access to quality medical equipment, overcrowding, unsanitary living conditions, and other similar characteristics. Some examples of locations where these countries are typically located include: Africa, the Caribbean, Latin America, and others (Ahmed, 2023). Detecting IPI presence in patients can be done through microscopy manually (visual inspection), but speeding up the classification process via Deep Learning can decrease the amount of time that the patients must wait prior to receiving the proper treatment.

**4. What type of data science task is it?**

This is a classification problem using supervised learning.

**5. Data: Brief description of data. How big do you expect the data will be? Is amount of**

**your data too big or too small? If you're web-scraping or collecting data, how long do you**

**expect to collect the data?**

**Dataset Link:** [**https://ieee-dataport.org/competitions/parasitic-egg-detection-and-classification-microscopic-images**](https://ieee-dataport.org/competitions/parasitic-egg-detection-and-classification-microscopic-images)

This dataset contains 11,000 images from microscopic analysis of samples from humans, all of which are split into 11 types of parasitic infections. I expect this dataset to be sufficiently sized to proceed with training of a neural network model, or similar. The data will not have to be web-scraped for collection, however there will be some cleaning to do to convert image files to usable input data to feed into CNN input layer(s).

**6. How will you analyze the data? What machine learning methods do you plan to use,**

**and/or what business intelligence aspect do you plan on incorporating?**

The plan is to analyze the data (images) by training different types of neural network models to identify patterns that can help classify the type of parasitic infection present. Before analyzing the data, I will have to clean the image file data by converting into pixelated values in a tensor/matrix format. Most likely I will build a convolutional neural network (CNN) for this project, and experiment with kernel sizes, loss functions, and other configurable parameters. Regarding libraries and existing tools for CNN modeling, I plan to take advantage of TensorFlow framework in Python and incorporate Transfer Learning to improve efficiency and, ideally, the model accuracy metrics.

**7. Describe any anticipated difficulties and problems. Discuss how you may overcome the**

**problems.**

I anticipate that there will be difficulty initially with getting a sufficiently high accuracy from the model, since there are 11 total labels for the data given. If this was a binary classification problem, for example, I would not be as concerned about achieving sufficient accuracy metrics. Regardless of if the dataset proves to be too small or not, I also intend to use data augmentation techniques for each parasitic egg type. This will allow me to increase the amount of data being processed in the CNN model training process. Another challenge is that the quality of microscopic images (brightness, resolution, etc.) may not be consistent across all the parasitic egg images. The model must be robust enough to handle these inconsistencies in the input image data.

**8. Suggest a timeline for the project. This should be a weekly breakdown of what you plan**

**on doing each week.**

**Week 2:** Data cleaning and sample image display to check for patterns

**Week 3:** Data augmentation and breaking into training, test, and validation splits

**Week 4:** Train simple CNN model(s) with varying layers, kernels, loss functions, etc.

**Week 5:** Transfer Learning to use portions of CNNs to improve computational efficiency

**Week 6:** Analyze/compare results from all models attempted and visualize/create accuracy metric plots for use in the presentation (confusion matrix demonstrating false positive/negative, etc.)

**Week 7-8:** Create, practice, and record presentation using PowerPoint

Note: Week 2/3/4 items could be combined to allow more time to explore and fine-tune CNN models to try for the highest accuracy possible

**9. Create GitHub repository for your Practicum project. Add this proposal, begin a ReadMe**

**document, and begin adding your data to your repository. Add a link to your GitHub**

**repository to this document**

**GitHub Link:** <https://github.com/LonnyCox55/MSDS692_ParasiticInfectionCNN/tree/main>

**References**

Bahrami, F., Haghighi, A., Khadem-Erfan M. B., Zamini, G. (2018). Diagnostic Accuracy of Intestinal Parasitic Infections in Individuals Admitted to Medical Laboratories. *Iran J Public Health*. 2018 Apr;47(4):620-622. PMID: 29900154; PMCID: PMC5996334. <https://doi.org/10.1038/s41598-023-41711-3>

Ahmed M. (2023). Intestinal Parasitic Infections in 2023. *Gastroenterology Res*. 2023 Jun;16(3):127-140. doi: 10.14740/gr1622. Epub 2023 Jun 11. PMID: 37351081; PMCID: PMC10284646.