The Potential of Artificial Intelligence to Identify Cancer and Aid Research

Author(s) Wilson B. Peguero Rosario

Master of Science in Data Science Capstone Design Specification

Grand Canyon University

Instructor: Professor TBA

Revision: 1

Date: December 7, 2022

**COPYRIGHTS**

MIT License

Copyright (c) 2022 Wilson Peguero Rosario

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

**Data Source**

Cui, Chunyan; Li Li; Cai, Hongmin; Fan, Zhihao; Zhang, Ling; Dan, Tingting; Li, Jiao; Wang, Jinghua. (2021) **The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast**. The Cancer Imaging Archive. DOI: <https://doi.org/10.7937/tcia.eqde-4b16>

**TABLE OF CONTENTS**

General Information 4

System Summary 5

Getting Started 5

Using the system 6

Troubleshooting 8

help and contact details 8

**General Information**

This is the guide to using the application known as “The Capstone Project”. This project may also be known as the tumor malignancy detector. This project has been fully written using python and libraries made by different communities. The libraries used to develop this application were:

* TensorFlow
  + Created by google
  + Used to develop machine learning models
* Pandas
  + Open-source library for working with tabular data
* Dash
  + Library for developing dashboards
* Pydicom
  + Library for reading DICOM files and extracting the data within said files.

The purpose of this project is to develop an API that allows users to create their own machine learning models and use them to make predictions on DICOM files.

**System Summary**

The system is comprised of an API for extracting, transforming, and loading DICOM data. This either loads singular DICOM files or large batches of DICOM files, allowing the user the ability to freely test, train, and use the API. The API comes together with a set of machine learning architectures for creating the machine learning models. Finally a standard manner of displaying the data through the use of the Dash library, yet the API provides users with the versatility to use the machine learning model to make predictions and display the data in any dashboard desired.

**Getting Started**

To begin using the modules, one can begin by organizing the data. The DICOM files may be organized in any fashion so long as they are referenced within an excel file containing the direct address to the DICOM files. In the case that the DICOM files do not contain the classification, please include the information within the reference file.

**Using The System**

Once the application has been fully setup, the user may be able to access the website. The website will appear like the image below:

Graphical user interface, application

Description automatically generated

Click on the drag and drop or Select Files area to load multiple images (alternatively, the images can be dragged to the same area.

A computer screen capture

Description automatically generated with medium confidence

Select the image(s) and click open. Some time may pass before the images are displayed, but the following is part of the dashboard after the images load:

Chart, bubble chart

Description automatically generated

Below is the second half of the dashboard:

Shape

Description automatically generated

The export button can then be used to download a copy of the table which will display as Data.csv.

**Troubleshooting**

Every function related to the API contains an error message in relation to its input and output. Python provides one with the ability to observe the exact location of the error together with the type of error that python has identified. Observe the page below for further information on the APIs: [Help | Python.org](https://www.python.org/about/help/)

For further troubleshooting of the dashboard or model, one can extract the csv file by following the instructions provided above and comparing the test data with the predictions displayed upon the dashboard and downloaded excel file.

**Help and Contact Details**

For further help, please file an issue at [New Issue · wpeguero/capstone (github.com)](https://github.com/wpeguero/capstone/issues/new). Please provide a description of the issue together with the error that you are currently encountering and the circumstances leading to the of the issue.