**Multi-Stage Cancer Classification**



Supervised By: TA: Yomna Ahmed

Designed By: Team SC\_2

Team Members

Ehab Mohamed Abdelmonem

Kerolos Helal

Mostafa Bayoumi

Habiba Kader

Noha Mohamed

Introduction:

As technologies expand, Machines showed their powers in almost all fields, they shorten time and help us make accurate decisions, we have been thinking of how machines can affect the medical field, well, Computer Vision, Machine Learning and similar techniques also showed very good results in those tasks, so here we will talk about how we used those techniques to implement a program that helps classifying scans and diagnosing them for cancer type.

Data Preparation and Preprocessing:

Data is consisted of images for scans, separated by scan type and diagnose, preprocessing workflow was as follows:

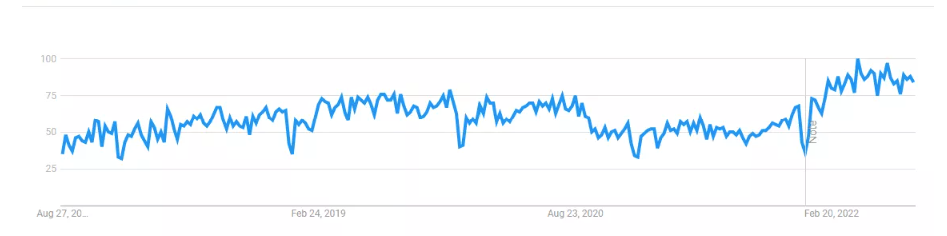
1. Created CSV file for all images in dataset with 3 columns:
2. path (image path)
3. type (scan type)
4. diagnose (cancer type)
5. Resized images to match needs:
6. 300\*300 for stage 1 (less features required)
7. 500\*500 for stage 2 brain class (high features scale)
8. 256\*256 for stage 2 breast class (CNN input size)
9. Color type (didn’t affect much)
10. Applied normalization at feature space
11. Train, Validation and Test split for breast data

Models and Techniques:

Stage 1:

* Feature Extraction: BOW with SIFT
* Classifier: Gradient Boosting (Ensemble Booster)

Gradient Boosting is a modern machine learning ensemble that uses a combination of classifiers, it evaluates itself by retraining misclassified data with different classifiers, figure below shows performance through time.



Stage 2 (Brain):

* Feature Extraction: HoG
* Classifier: AdaBoost (Ensemble Booster)

AdaBoost is an ensemble that reassigns weights to instances of data but with higher weights for misclassified cases.

Stage 2 (Breast):

* CNN: Higher Accuracy
* AlexNet CNN: More complex
* MLP Classifier using HoG: Lower accuracy but stable

Time Analysis:

|  |  |
| --- | --- |
| **Method** | **Time** |
| BOW | 7 min + 22 min for KMeans |
| Gradient Boosting | 3s for Train, 1s for Test |
| HoG | 117s |
| AdaBoost | 50s for Train, 1s for Test |
| Custom CNN | 185s for Train, 4s for Test |
| AlexNet CNN | ~7 min for Train, 4s for Test |
| MLP (Breast) | 16s for Train, 3s for Test |

Accuracy Analysis:

|  |  |  |
| --- | --- | --- |
| **Method** | **Train** | **Test** |
| Gradient Boosting | 1.0 | 0.977 |
| AdaBoost | 1.0 | 0.965 |
| Custom CNN | 0.915 | 0.91 |
| AlexNet CNN | 0.883 | 0.744 |
| MLP (Breast) | 0.99 | 0.68 |

Visualization:

Chart

Description automatically generated

Feature Visualization after applying SIFT

A picture containing text

Description automatically generated

Feature Visualization after applying HoG for Brain

Text

Description automatically generated with medium confidence

CNN Training Process

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Loss and Accuracy Visualization

Tools:

Jupyter Notebook, Colab, PyCharm