ACTIVE LEARNING PROJECT 2 TECHNICAL REPORT

Exploring Active Learning Strategies and Classifiers on Different Datasets

Implemented By:

Arwa Sallam	20200067
Natalie Monged	20200586
Mahmoud Wael	20200505
Aly Walid	20200336
Lilian Stevens	20200404

MAY 2, 2024

Abstract:

Active machine learning is a powerful approach that iteratively selects the most informative data points for labeling, thereby reducing the labeling effort while maintaining or improving the model's performance. Active machine learning addresses this challenge by actively selecting the most informative instances for labeling, thereby maximizing the performance gain with minimal labeling effort. In this report, we present our implementation of active machine learning on various datasets using a combination of classifiers and uncertainty sampling strategies.

Introduction:

Conventional supervised learning requires large datasets with supervised labels to form perfect models. This involves labelling many data sets which can be costly and time-consuming. In this technical report, we implement active machine learning strategies on two different image datasets which are: CIFAR-10 and Flowers Recognition using multiple classifiers. The classifiers in our implementation include Support Vector Classifier (SVC), Random Forest and a Neural Network classifier using the keras classifier. The uncertainty sampling strategies include entropy sampling, margin sampling and query by committee using vote entropy. We aim to show through various active learning experiments that we can achieve higher metrics using deep learning models while lessening the amount of labeled data needed.

Implementation Design:

- **Datasets Selection:** We utilized two widely used datasets for classification tasks: CIFAR-10 and Flowers Recognition

- **Classifier Selection:** We employed multiple classifiers to evaluate the performance of active learning. These classifiers include Support Vector Classifier (SVC), Random Forest (RF), and Neural Network (NN).

- **Uncertainty Sampling (US) Strategies:** We applied uncertainty sampling strategies to select the most informative instances for labeling. The strategies include Entropy sampling (ES), Margin Sampling (MS), Random Sampling (RS) and Query-By-Committee (QBC) using Vote Entropy.

- **Active Learning Cycle:** We implemented an active learning loop that iteratively selects instances for labeling based on the chosen sampling strategies. After each iteration, the selected instances are labeled, and the classifier is retrained on the updated labeled dataset.

- **Evaluation Criteria:** We applied metrics like F1 Score, Confusion Matrix and Accuracy which were sufficient for comparison as shown below in the next section.

Datasets:

As mentioned before, we worked on two distinct datasets: CIFAR-10 and Flowers Recognition. In the subsequent sections, we will delve into each dataset individually, discussing the methodologies employed and the outcomes of the various experiments conducted.

- Flowers Recognition:

This dataset comprises 4242 images of flowers categorized into 5 classes: Chamomile, Tulip, Rose, Sunflower, and Dandelion.

- **Model Comparison:** We evaluated the performance of various deep learning CNN models on the Flowers Recognition dataset. The models were tested and their corresponding testing accuracies are as follows:

Model	Testing Accuracy
MobileNet	90.7%
ResNet50	45%
EfficientNetB0	25.9%
InceptionV3	86%
VGG19	72.6%

  As shown in the table, the MobileNet model demonstrated the highest performance among the tested models, so we decided to use it for further refinement.

- **Activation Function Comparison:** To further optimize the MobileNet model, we experimented with different activation functions. The architecture used consisted of the MobileNet model followed by a Global Average Pooling Layer, a Dense Layer with 128 Neurons, and finally an output classification layer with 5 neurons. We compared the testing accuracies obtained using three different activation functions:

Activation Function	Testing Accuracy
ReLU	91.6%
Tanh	91.4%
Sigmoid	90.7%

Although the results were close, the Tanh activation function was selected for further experimentation.

- **Active Learning Experiments:** To further enhance the model's performance, we conducted active learning experiments by varying the size of the initial data and employing different query strategies. The details of these experiments are outlined below:

Initial Data Size	No. of Queries / Loops	Instances per Queries / Loops	Query Strategy	Testing Accuracy
341 (1/10 of the training data)	10	5	Entropy Sampling	87%
682 (1/5 of the training data)	10	5	Entropy Sampling	88.5%
852 (1/3 of the training data)	10	5	Entropy Sampling	88%
1410	10	5	Entropy Sampling	89%

After careful consideration of various data portion sizes, query or loop counts, and instances per query loop, we concluded that the configuration comprising an initial data size of 682 (equivalent to 1/5 of the training data), 10 queries / loops, and 5 instances per query / loop yielded the most promising results. Therefore, we determined this setting to be the most suitable for our experimentation moving forward.

- **Results without Active Learning:** For reference, without employing active learning, we trained the MobileNet model using 80% of the data for training, with 30% of the training data reserved for validation. The model was trained for 10 epochs with a batch size of 200, resulting in a testing accuracy of 91% and F1 Score of 86%.

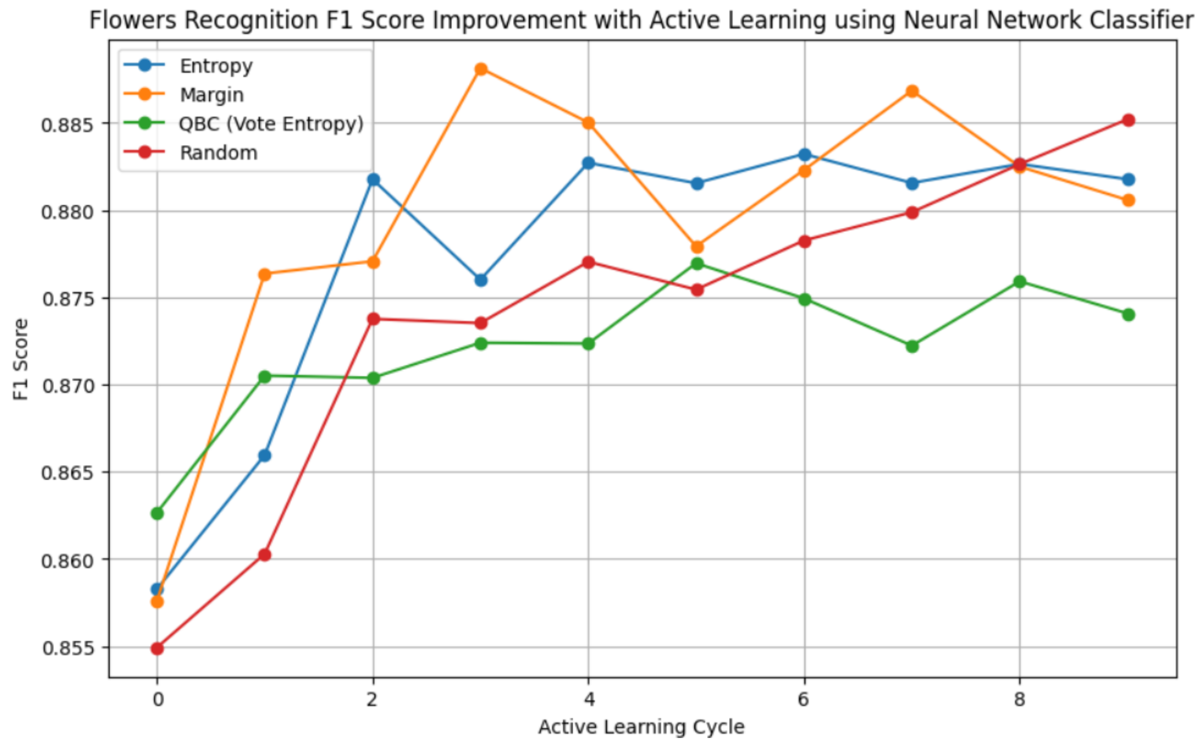
- **Classifier Comparison after Feature Extraction:** Following feature extraction using the MobileNet Model followed by the architecture mentioned before, we applied three different classifiers with Entropy Sampling and evaluated their performance based on the F1 score:

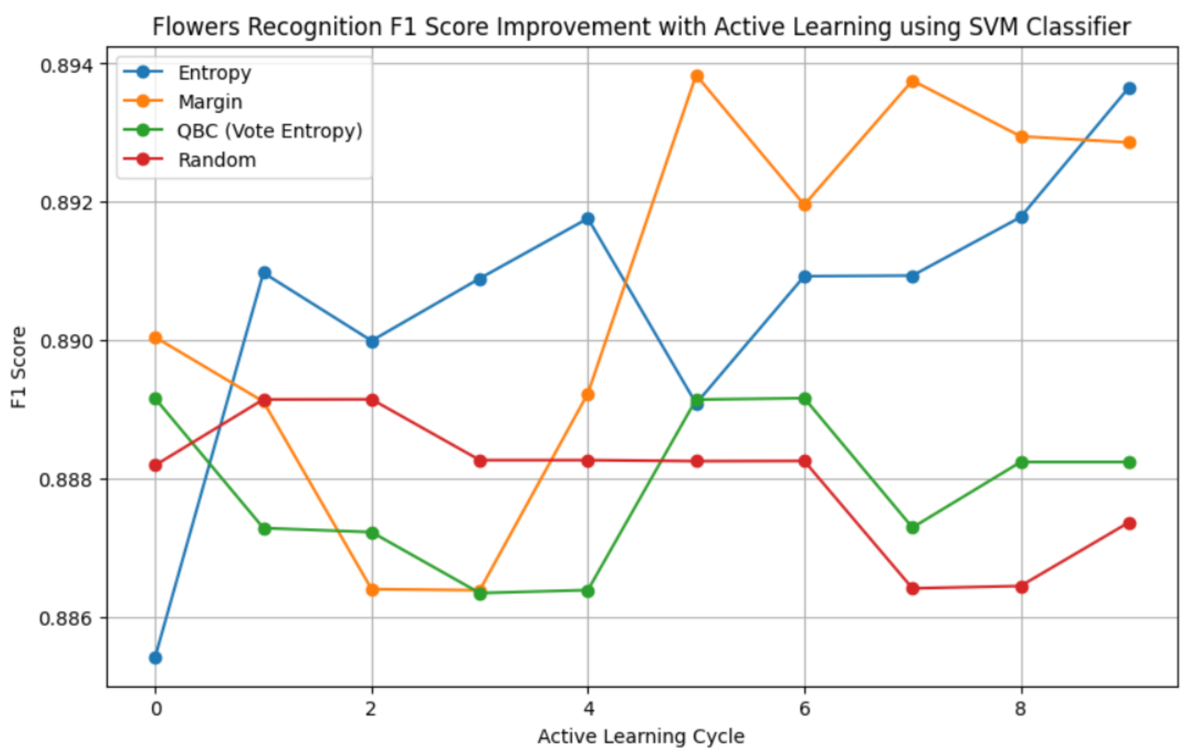
Classifier	F1 Score
Support Vector Machine (SVM)	88%
Random Forest Classifier	84%
Neural Network Classifier (Softmax)	86%

- **Query Strategy Comparison:** We explored four distinct querying strategies: Entropy Sampling, Random Sampling, Query-By-Committee (Vote Entropy), and Margin Sampling. Each strategy was evaluated on both the Support Vector Machine (SVM) classifier and the Neural Network (NN) classifier using the F1 Score on the test data.

Strategy / Classifier	Random Sampling	Margin Sampling	Entropy Sampling	Query-By-Committee (Vote Entropy)
SVM	88%	89%	88%	88%
Neural Network Classifier	88%	88%	87%	87%

The F1 Scores Improvement plots:

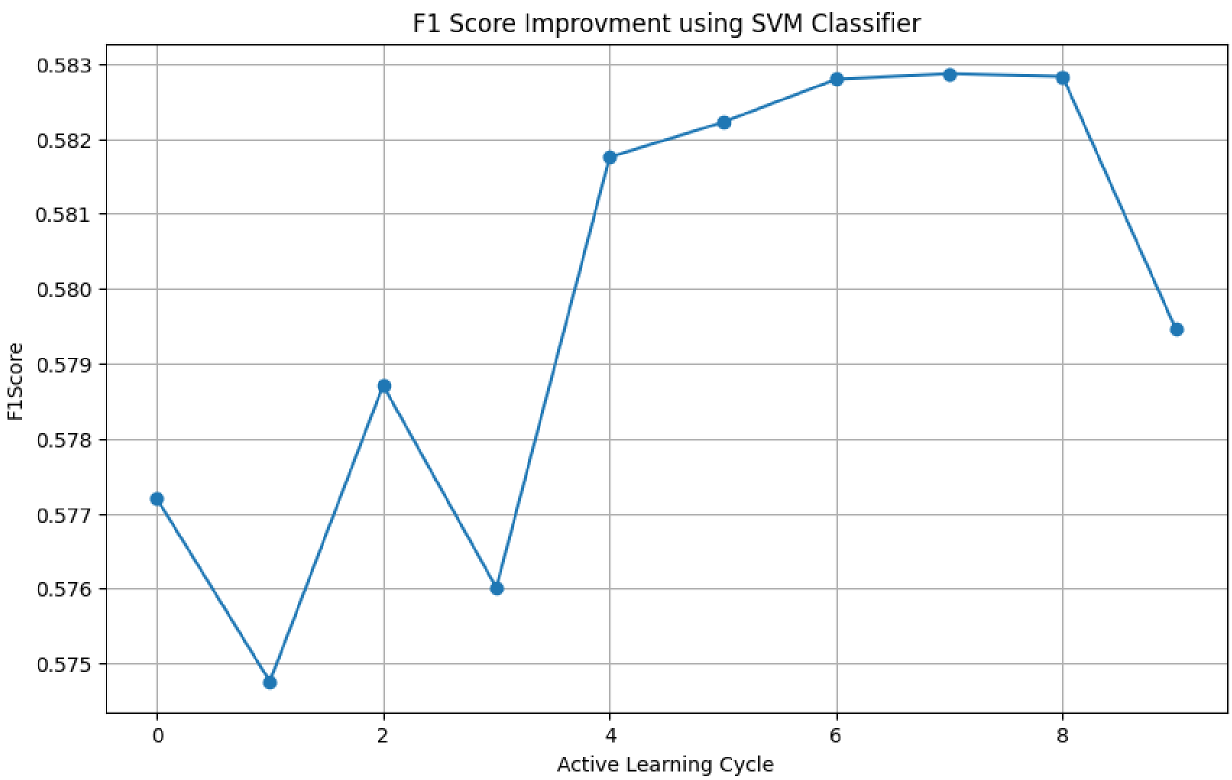




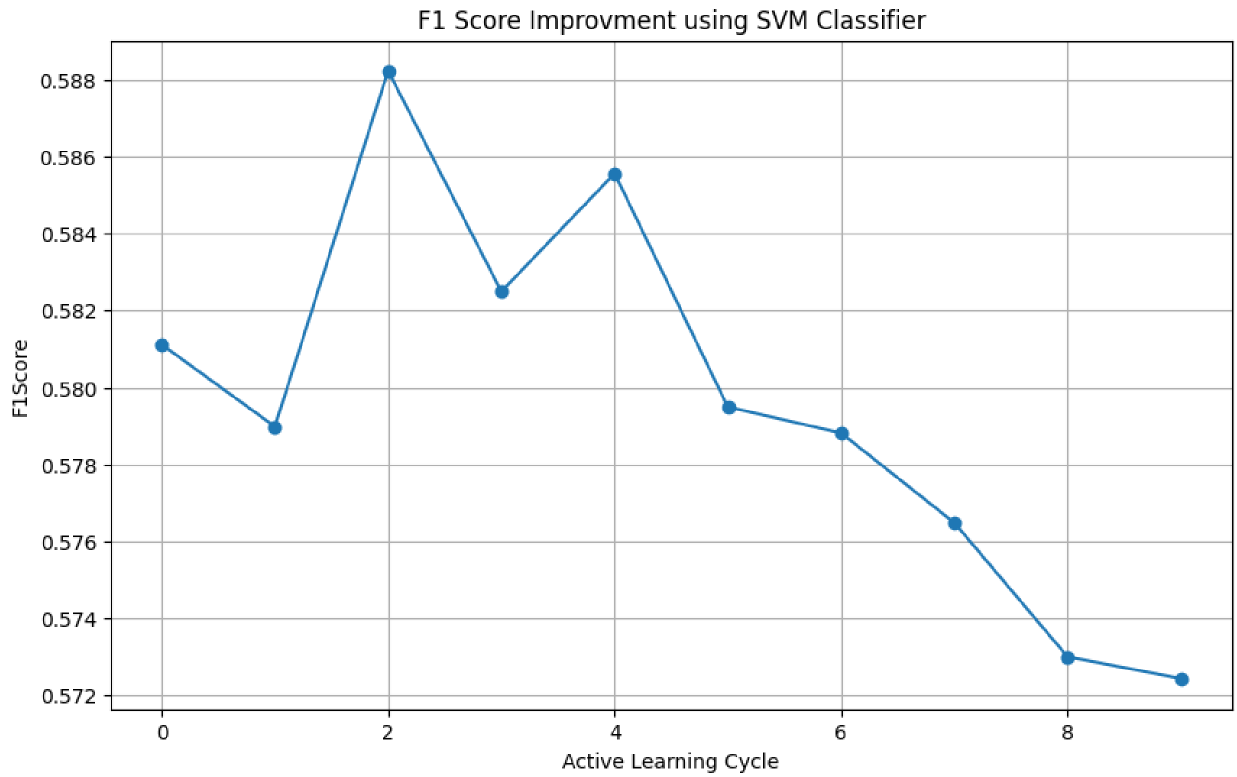
- **SVM Classifier Performance Visualization:** To visualize the performance of the SVM classifier throughout the active learning cycles and observe the behavior of the data distribution as new instances are queried, Principal Component Analysis (PCA) was employed to reduce the feature dimensionality for plotting. The classifier was tested on two different portions of the data using the entropy sampling strategy.

Initial Data Size	No. of Queries / Loops	Instances per Queries / Loops	F1 Score on test data	Accuracy on test data
682	10	5	20%	20%
658	10	150	17.7%	18%

The F1 Scores improvement plot on initial data size of 682:



The F1 Scores improvement plot on initial data size of 658:



Conclusion:

In this project, we explored two datasets, CIFAR-10 and Flowers Recognition, employing various deep learning models, activation functions, active learning strategies, and classifiers.

- **Flowers Recognition:**

  As for the Flowers Recognition dataset, we identified through rigorous experimentation that the MobileNet model is the most effective for flower recognition, achieving a testing accuracy of 90.7%. Additionally, The Support Vector Machine (SVM) emerged as the top performer, achieving an F1 score of 88% on testing data. Notably, the F1 scores achieved by different strategies were very close to each other, yet the Margin Sampling strategy achieved the highest F1 score of 89% on test data.
  Active learning experiments highlighted the importance of selecting appropriate initial data sizes and query strategies, with promising results observed for a combination of 682 initial data size, 10 queries or loops, and 5 instances per query loop. Visualization of the SVM classifier's performance throughout active learning cycles provided insights into data distribution changes. Our findings underscore the significance of thoughtful model selection, activation function optimization, and active learning strategy in achieving robust performance for

flower recognition tasks, with implications for future machine learning endeavors in similar domains.