Simple Evolutionary Exploration into Classification Algorithms

Davin Lin[1]; Dirk Colbry, PhD[2]
[1]Grinnell College, [2]Michigan State University

Background

- Genetic algorithms (GAs) find good solutions in large solution spaces to search problems through a process inspired by evolution.
- There have been several studies that use GAs to search over hyper-parameters of machine learning algorithms to learn values that work well for specific problems.
- The SEE Toolkit which uses and implements a Simple Genetic Algorithm **(Figure 1)** to support research workflows does not support searching over supervised-learning machine learning algorithms.

Goals

- To **extend** the SEE toolkit and framework to **support** supervised-machine learning classification algorithms
- To **benchmark** its performance via a comparison with a **previous work** on hyperparameter tuning via genetic search

Methods and Materials

Methods:
1. 30 trials of GA for 100 iterations and population of 100
2. Fitness Function: **(Figure 2)**

$$Error\ Rate = \frac{number\ of\ incorrect}{number\ of\ classifications}$$

Materials:
1. Breast Cancer Wisconsin (Diagnostic) Dataset
2. Dhahri et al. 2019 (study on hyperparameter tuninng)
3. High Performance Computing Center at MSU ICER
4. Scikit-learn package for machine learning

Figure 1. Simple Genetic Algorithm

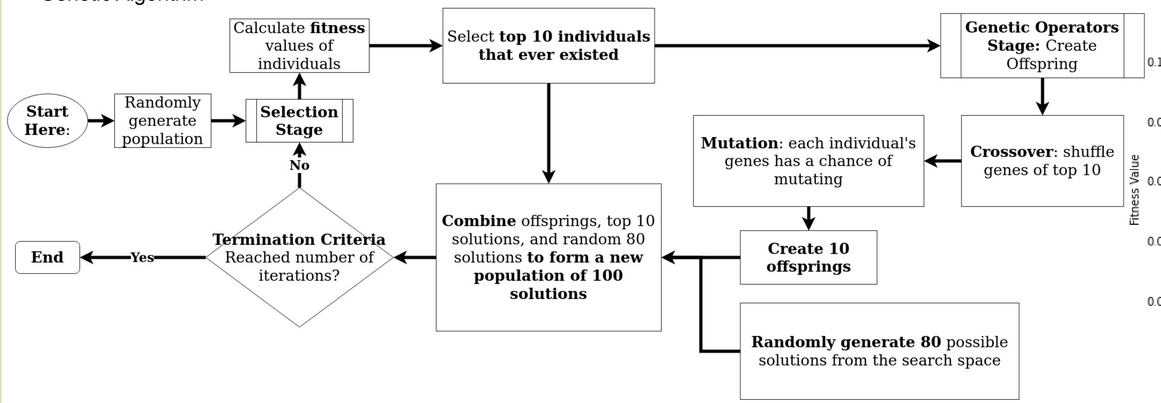


Figure 2.
Fitness Function

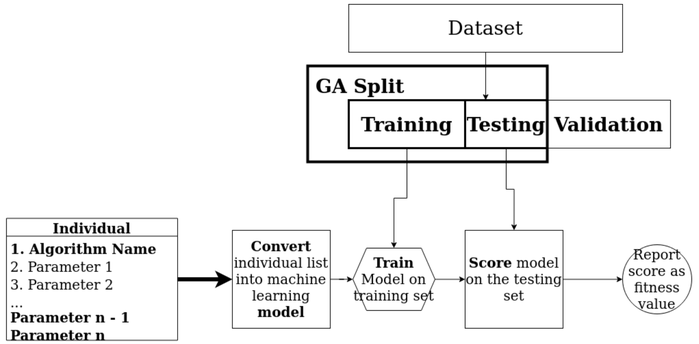
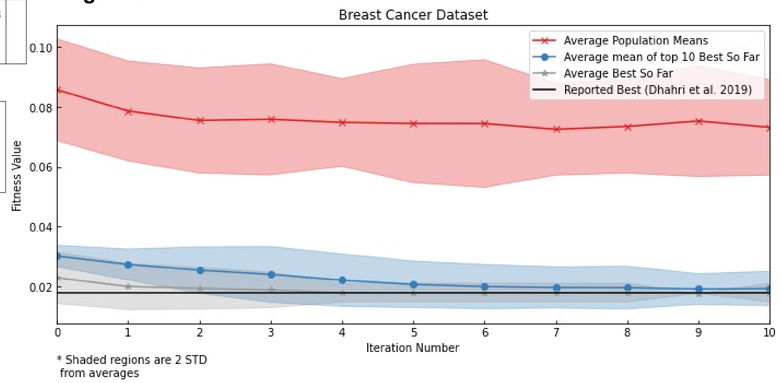


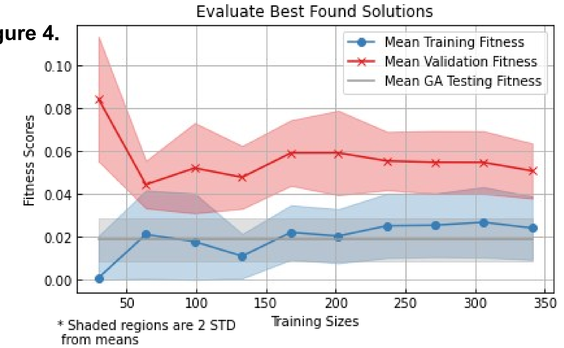
Figure 3.



Results*

- Our framework is able to find solutions similar to reported results **(Figure 3)**
- While the best found solutions converge during GA, when evaluated independently on the validation set, their fitness scores fall far from the range of performance during GA **(Figure 4)**.

Figure 4.



Conclusions and Future Work

We extended the SEE framework to support classification algorithms and benchmarked the performance of its simple evolutionary approach.

Future Work:
- The implemented simple GA is a **naive** approach and can be greatly improved.
- Benchmark GA on alternate real datasets and using more complex metrics for the fitness function

MICHIGAN STATE UNIVERSITY
CMSE
Grinnell College