Optimizing Machine Learning Based Oil Spill Detection Using Genetic Algorithm Techniques and Adaptive Chromosome Replacement (ACR)

Justin Yu, Los Altos High School, Los Altos, CA | Wesley Luh, Leland High School, San Jose, CA

Introduction

Ecological Impact

The most pressing effects of oil spills and discharges are animal injury or mortality. Consumption of petroleum oil has been linked to issues in the renal, hematologic, and neurological systems. As these issues manifest in humans, psychological and socioeconomic issues arise. Because petroleum oil most heavily affects low level feeders, the process of biomagnification leads to these toxins affecting high level feeders, such as humans, to be harmed.

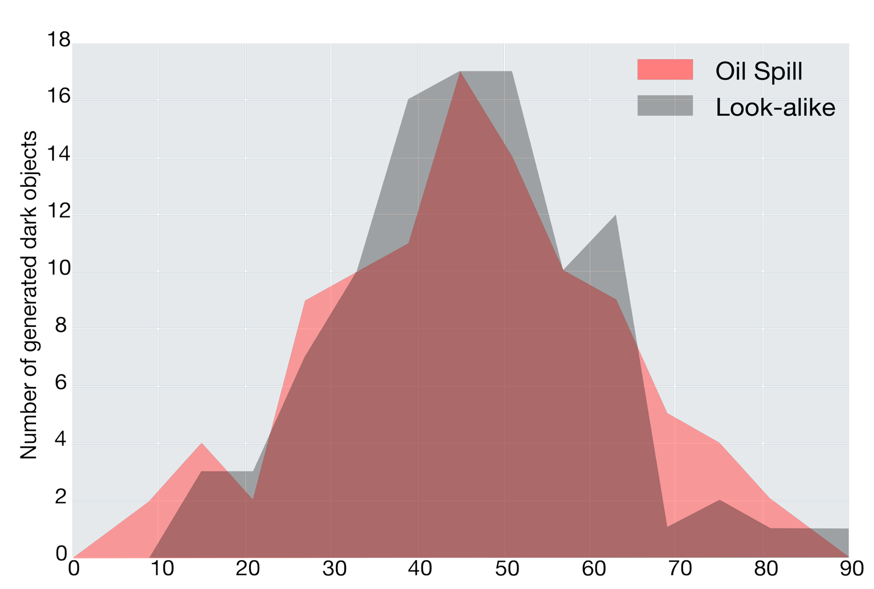
Remote Sensing

Remote sensing is the application of high flying satellites and sensors to acquire data. One form of remote sensing used in this study, the Synthetic Aperture Radar (SAR) imagery system, collects data by emitting an electromagnetic frequency and then combining signals that reflect off the Earth's surface. In the context of oil spill detection, potential oil spills show up as dark spots on SAR images, though these dark spots can also represent lookalike oceanic films.

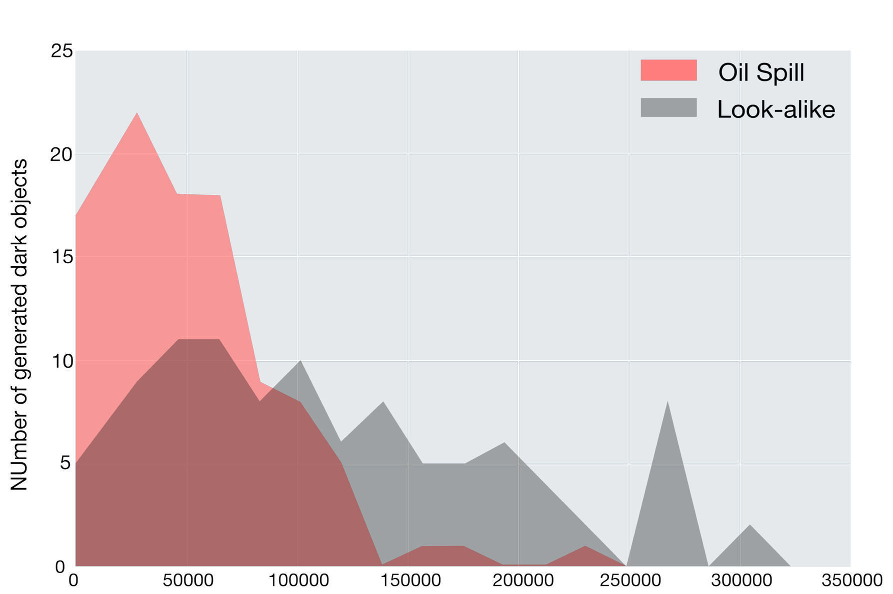
Genetic Algorithm

In previous studies, the accuracy of the traditional machine learning algorithms is limited by the presence of look-alike films that have overlapping physical features with true oil spills, since both show up as dark spots in SAR images. The classifier takes in a set of input features and outputs either a 0 (oil spill) or a 1 (lookalike). We propose a genetic algorithm to address the challenge of identifying useful features from data. The genetic algorithm models the biological process of natural selection to remove unnecessary features and increase classification accuracy.

Data & Methodology



Distribution for the mean of object backscatter intensity values (in decibels) in dark objects.



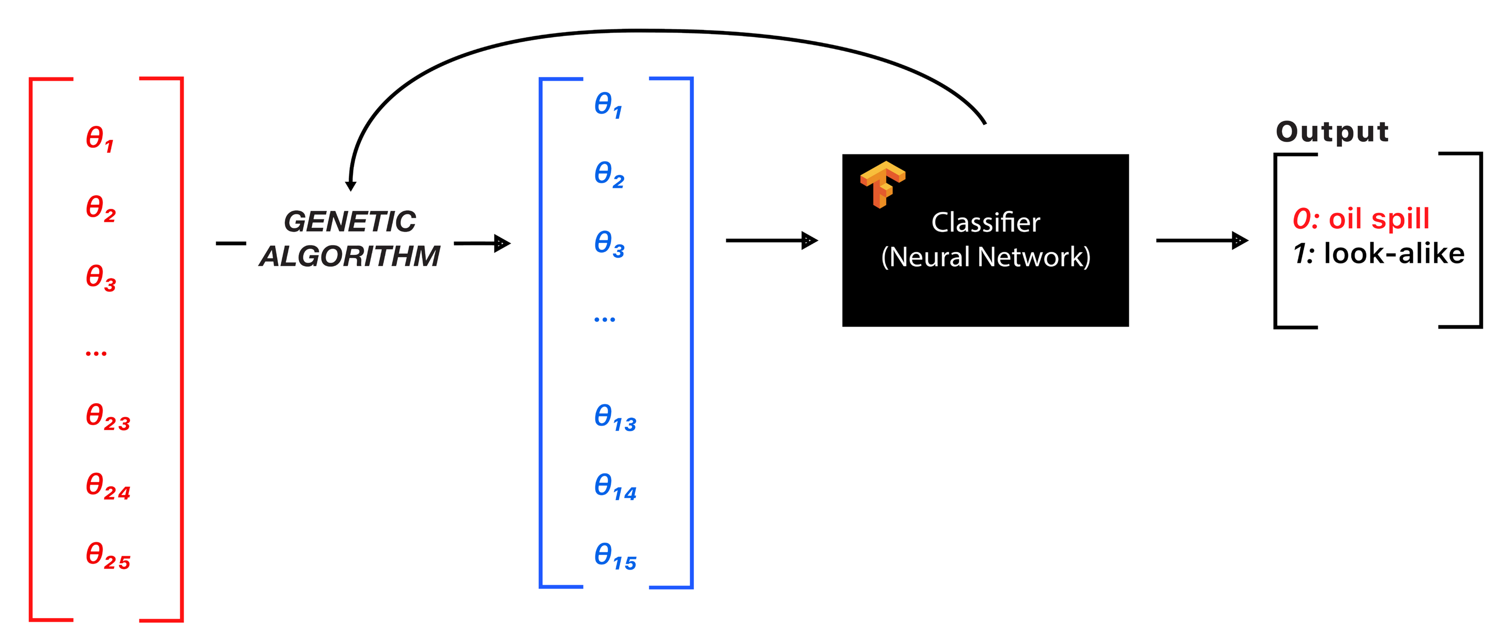
Distribution for the areas of detected dark objects (in m²), with oil spills red and look-alike objects in grey.

Question

How will the implementation of ACR to a genetic algorithm affect the accuracy of dark object classification?

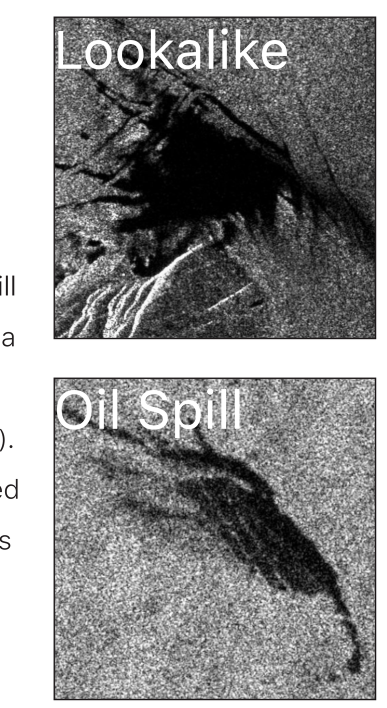
Hypothesis

If ACR is introduced to a genetic algorithm used to optimize the classification of dark objects from SAR satellite data into oil spills and look-alike films, then the accuracy will compare favorably to traditional methods of machine learning.



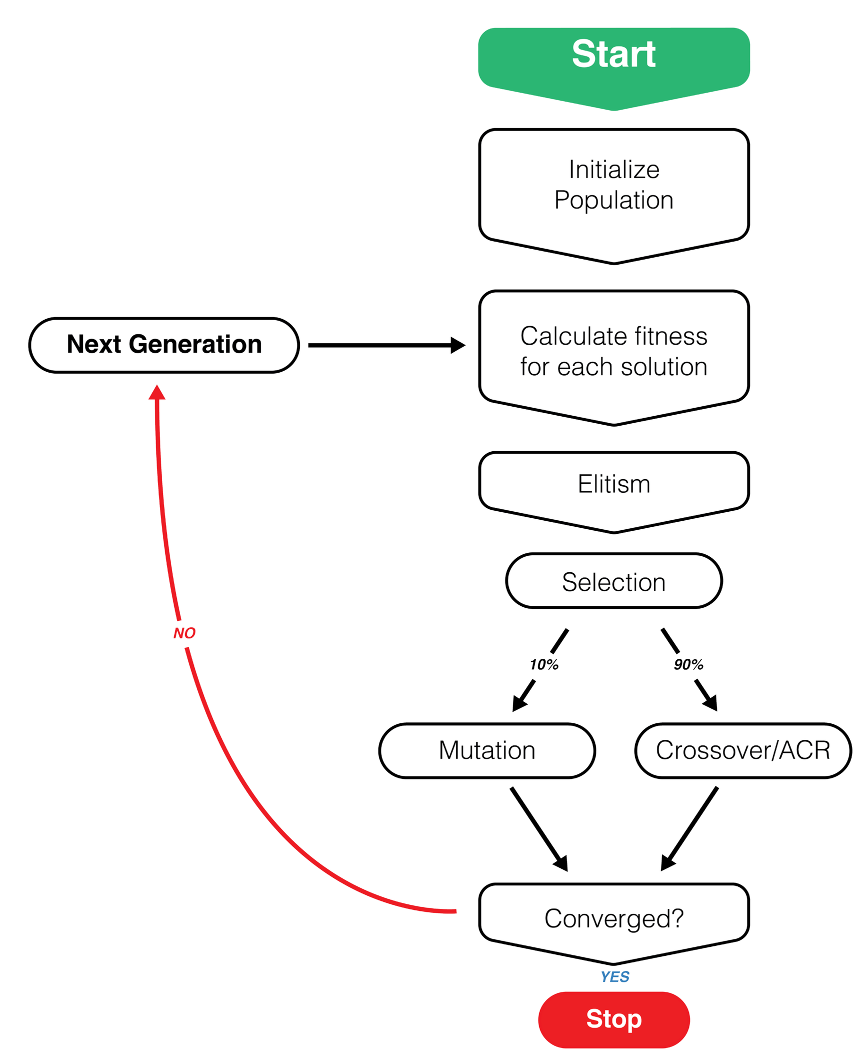
Right: The mean classification accuracy of the population of solutions throughout the course of one trial (50 generations): ranges from around 69% to 81% in overall accuracy for an improvement of 12% accuracy from the original population.

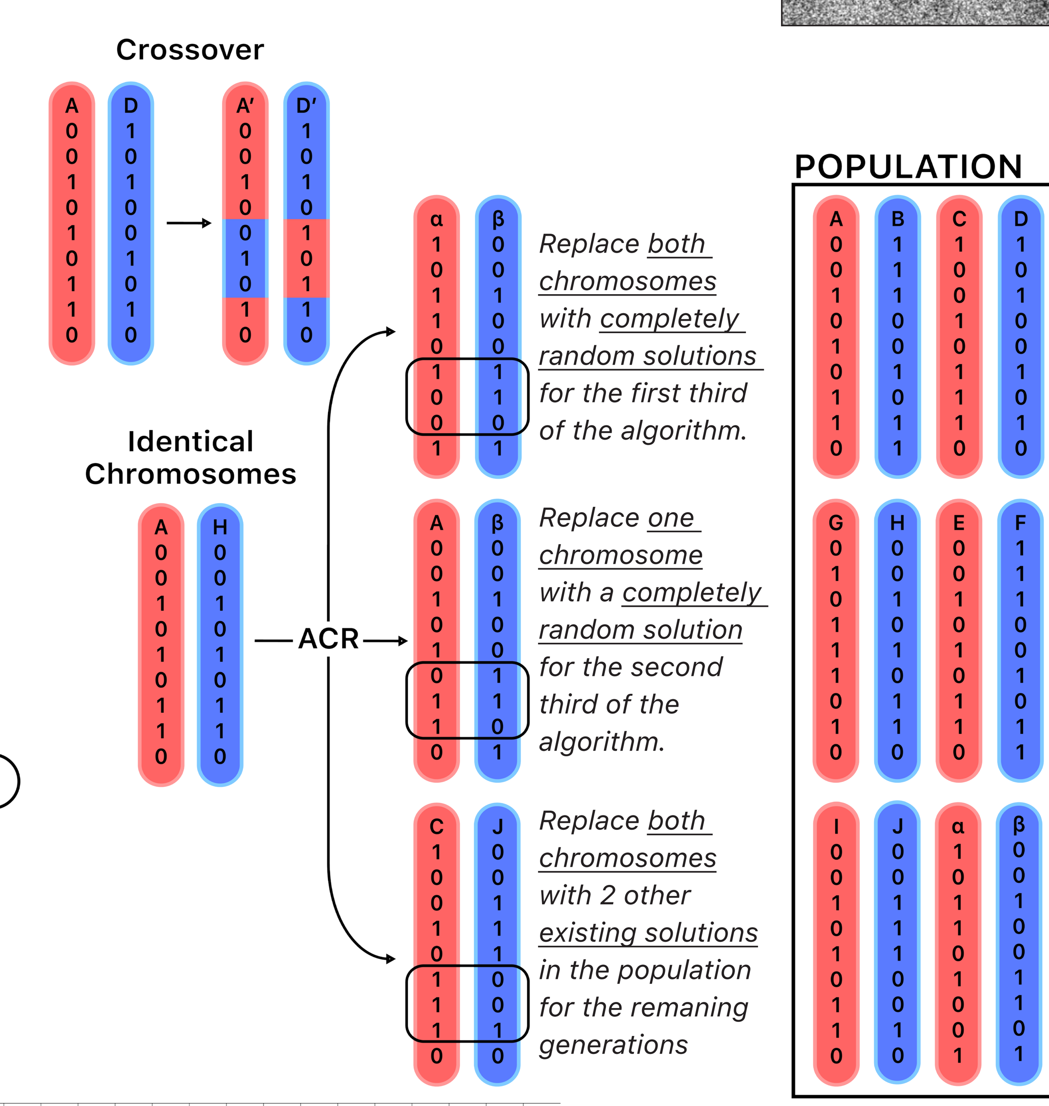
Left: Comparison between a verified oil spill (bottom) and a verified look-alike film (top). Figure adapted from Stathakis et al. (2006).

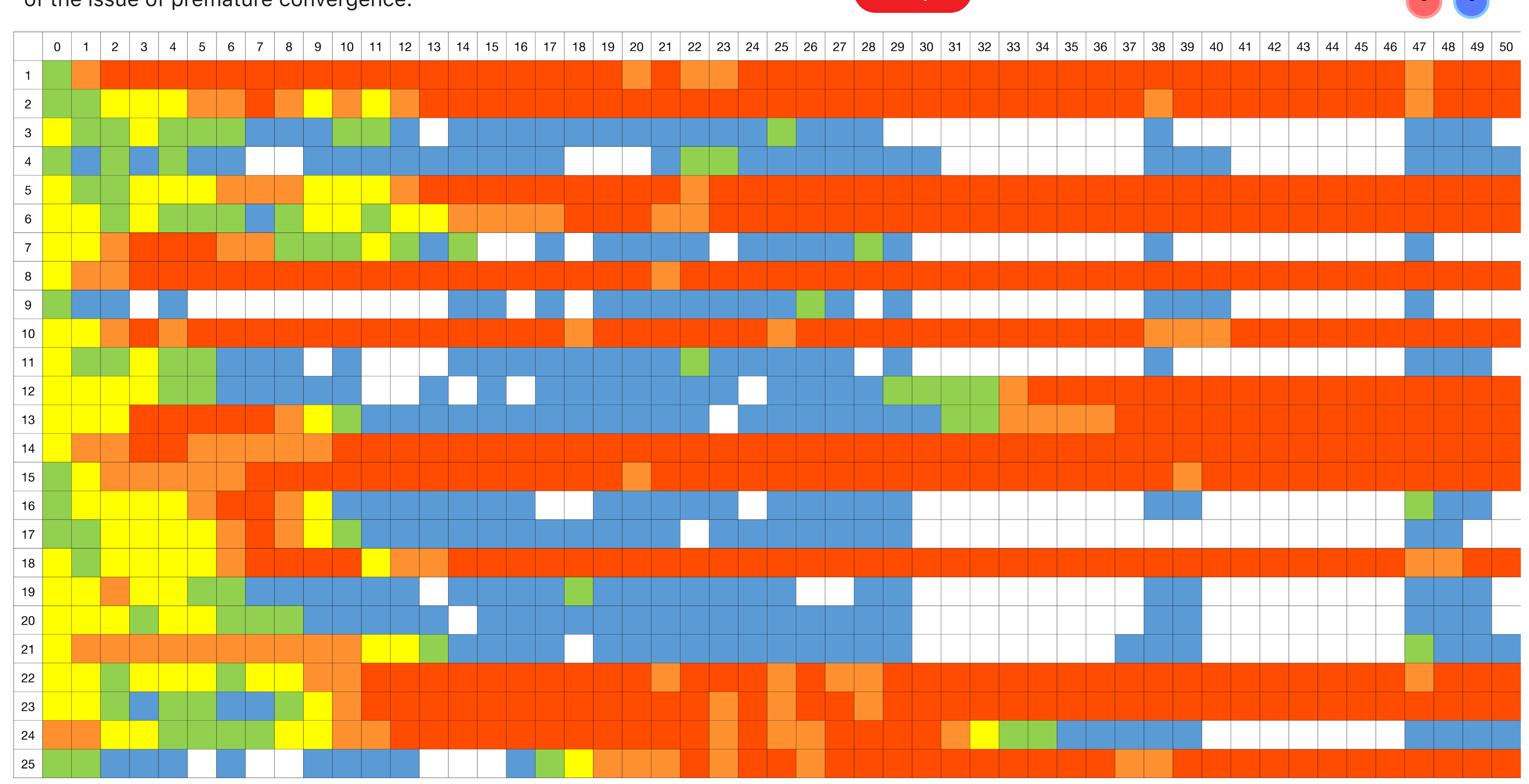


Above: This flowchart diagrams the process of isolating the feature with the most statistical impact. Initialization occurs with the values for each feature. The genetic algorithm (GA) reduces the feature count to optimize the signal-to-noise ratio (see above for detail). This reduced set of features is passed through the classifier, which returns the accuracy of the featureset, which is then further processed by the GA. After 50 generations, a final output accuracy is given.

Right: A genetic algorithm mimicks the biological process of natural selection in the real world. ACR is an additional process appended to the process of genetic crossover that is applied selectively when the information is identical. This allows for the prevention of the issue of premature convergence.







x = frequency (%) that feature K is used in a given generation of 20 solutions

White:	$x = 0$
Blue:	$0 < x \le 20$
Green:	$20 < x \le 40$
Yellow	$40 < x \le 60$
Orange:	$60 < x \le 80$
Red	$80 < x \le 100$

The best solutions recorded from 15 different trials of the genetic algorithm. A highlighted '1' means that the feature at that index is used in classification. (see table to the right).
Notably, the first and tenth features (area and background mean intensity) were used in every solution. The average true positive rate (oil spill detection accuracy) is 88.2%.

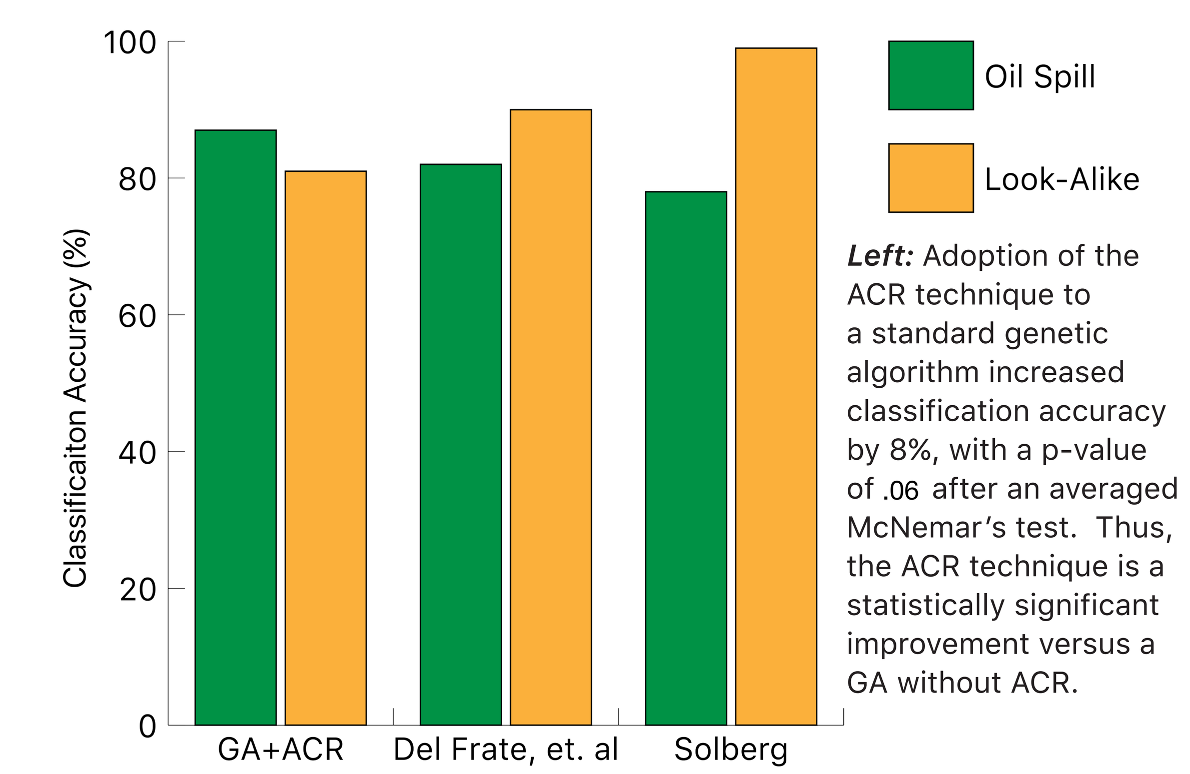
Discussion

ACR's improvement of 9% over previous GA attempts is undoubtely a viable technology applicable to future work involving GA. Application of the ACR technique can be used in other fields requiring the classification of different types of lookalike objects, spanning across fields including agriculture and biological research.

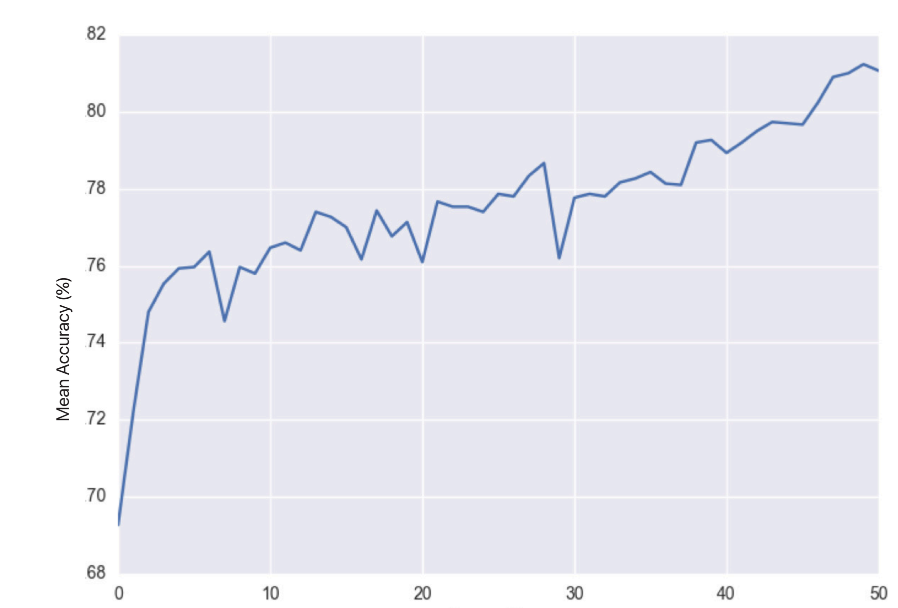
Limitations

The main limitation of this study was the lack of true, labeled oil spill data. Thus, data was generated based off of published statistics adapted from Toupoulezis and Solberg.

Conclusion



Left: Adoption of the ACR technique to a standard genetic algorithm increased classification accuracy by 8%, with a p-value of .06 after an averaged McNemar's test. Thus, the ACR technique is a statistically significant improvement versus a GA without ACR.



Left: The mean classification accuracy of the population of solutions throughout the course of one trial (50 generations): ranges from around 69% to 81% in overall accuracy for an improvement of 12% accuracy from the original population.

Future Work

Gaining access to confirmed satellite data of oil spills and lookalikes would also allow the system to work in a more realistic test environment. Accuracy measurements in this study may have been artificially specialized to certain datasets. Future applications of the ACR technique include applications to other fields, such as topology and biology.

Acknowledgements

We are indebted to the National Aerospace and Space Agency for providing us with the means to present the research study today. We would like to thank Dr. Weile Wang for mentoring and providing us with his insight and experience to our research, as well as giving us feedback on our presentation. We extend our thanks to our family members for supporting us along the way. We thank Ms. Karen Davis for sponsoring this project.