Optimizing Automatic Oil Spill Detection using Genetic Algorithm Techniques

Abstract. Oceanic oil spill pollution poses a serious threat to marine ecosystems. Undetected spills from accidents or operational discharges from ships can harm marine life residing on the sea surface or on shore. Previous studies try to automatically detect spills using remote sensing imagery from synthetic aperture radars. However, the accuracy of the traditional algorithms is limited by the presence of look-alike biogenic films that have overlapping physical features with true oil spills. Difficulties remain in identifying useful features and their combinations to distinguish between the two categories. Here, we proposed a methodology based on a genetic algorithm to address this challenge. This procedure finds an optimal or near-optimal solution within a given search space by mimicking the natural processes of competition, reproduction, mutation etc. Our procedure also addresses the premature convergence problem commonly found in genetic algorithms. We tested the procedure on simulated datasets using published statistics, and our results show an average accuracy of 88.2% for oil spills, representing a statistically significant improvement over the literature (~82%). The optimized classification also reduces the time spent on look-alikes, allowing for a faster response to minimize environmental damage.

1 Introduction

1.1 Impacts of Oil Spills

One of the most pressing issues regarding oil spills is animal injury or mortality. This comes in many forms, including the consumption of oil-contaminated shellfish by sea otters and the death of seabirds when coated in an oil film. As demonstrated in the Exxon-Valdez oil spill of 1989, shellfish absorbed the oil when residing in contaminated beds of water. Sea otters that later consumed these contaminated shellfish were harmed by the petroleum. The consumption of petroleum has been linked to issues in the renal, hematologic, and neurological systems [2].

Sea birds dive into the surface of the water to find food, but as they plunge past the petroleum layer, the birds get coated in a layer of oil, which is distributed over the bird's plumage. As temperatures fall, the oil inhibits the plumage's insulation and can lead to hypothermia or even death. In Spain, similar symptoms and environmental impacts were present five years after a major oil spill. Overall, both small and large oil spills can have devastating consequences that travel from the bottom of the food chain all the way up to human beings [4].

1.2 Remote Sensing

Remote sensing (RS), the act of acquiring data pertinent to a certain object or area without physically touching the object, is accomplished using satellites or other high-flying aircraft instruments. There are two major categories of remote sensing: passive and active sensing. Passive remote sensing measures wavelengths from the Sun that reflect off the Earth's surface, whereas active sensors produce their own frequencies to measure reflectance [15].

Radar is a form of active sensing that emits electro-magnetic waves to identify and map objects. It is commonly used for oil spill detection and many similar applications because radar data is consistent and not significantly affected by time of day, cloud cover or weather conditions, which is important for routine oil spill surveillance [6].

1.3 Synthetic Aperture Radar (SAR)

SAR is a method of remote sensing captures high resolution 2-D maps of a certain topographical area with a relatively small antenna. SAR works by combining data collected from multiple positions in space to create a synthetic image. The combination of images is also the reason why SAR images inherently contain “speckle noise” [3].

Satellites that collect SAR data include the Sentinel-1, RADARSAT-1, RADARSAT-2, ENVISAT ASAR and SEASAT missions. Each SAR instrument can send out unique frequencies, referred to as their band. For example, NASA’s SEASAT satellite used an L-band system (with wavelength of 15-30 cm). The RADARSAT-1 and RADARSAT-2 missions currently use a C-band system (with wavelength of 3.75-7.5 cm) to observe the ocean surface [12]. Along with band distinctions, SAR satellites also differ by their mode and their spatial resolution. A satellite’s spatial resolution measures the scaled width of each pixel in the received image. In this study, SAR data is in the form of backscatter values, calculated from the signal that reflects off the Earth’s surface. After calibration, backscatter values are transformed into data in decibels (dB).

Polarization is used by satellites to measure magnitudes of reflection as data. Satellites measure the difference between the multiple types of polarizations to determine the reflectance of a certain area. A radar system can have up to four channels of polarization: HH, VV, HV, and VH (H refers to horizontal, V refers to vertical). The first letter signifies the transmission polarization; the second letter signifies the receiving polarization. HH and VV are like-polarized because the transmission and receiving polarization is the same. HV and VH are cross-polarized because the transmission and receiving polarizations are perpendicular. Radar systems have the ability to use one, two, or four polarizations at once [16].

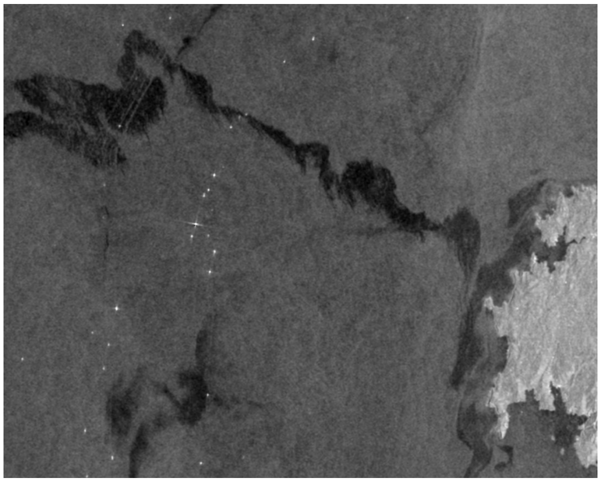

Figure 1. SAR image of the 2002 Prestige oil spill in Galicia, Spain after standard calibration and speckle filter processes (data preprocessing methods). Image generated from sample data provided by the European Space Agency.

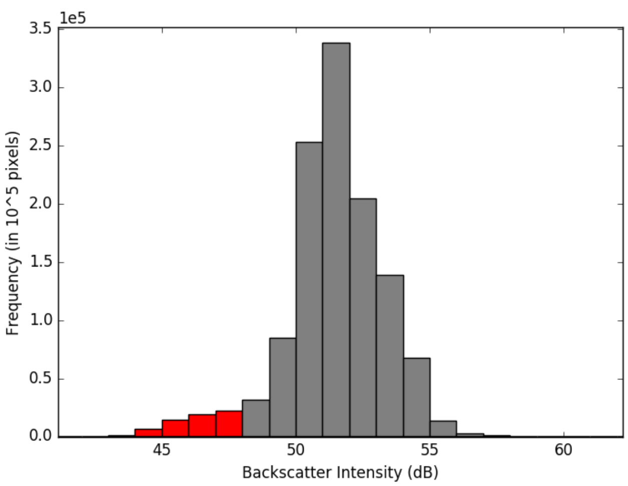

Figure 2. Distribution of backscatter values (mean = 50.15, std. deviation = 1.49) for the spill in Figure 1. The red represents the pixels with intensities less than the threshold (47.17).

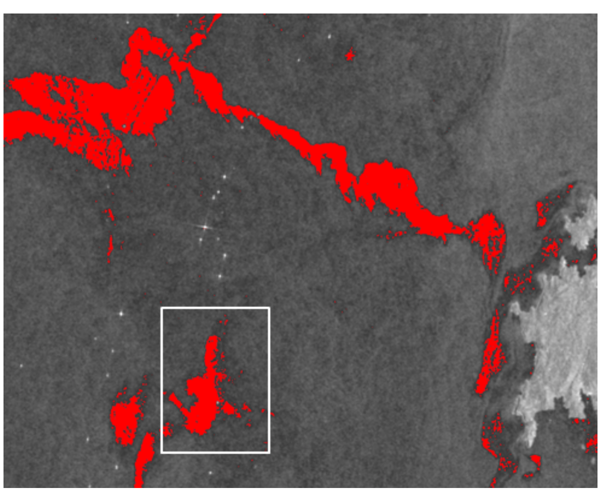

Figure 3. Identified dark spots highlighted in red to correspond with the red histogram values in Figure 2. The boxed spill is an example of what a singular dark object looks like.

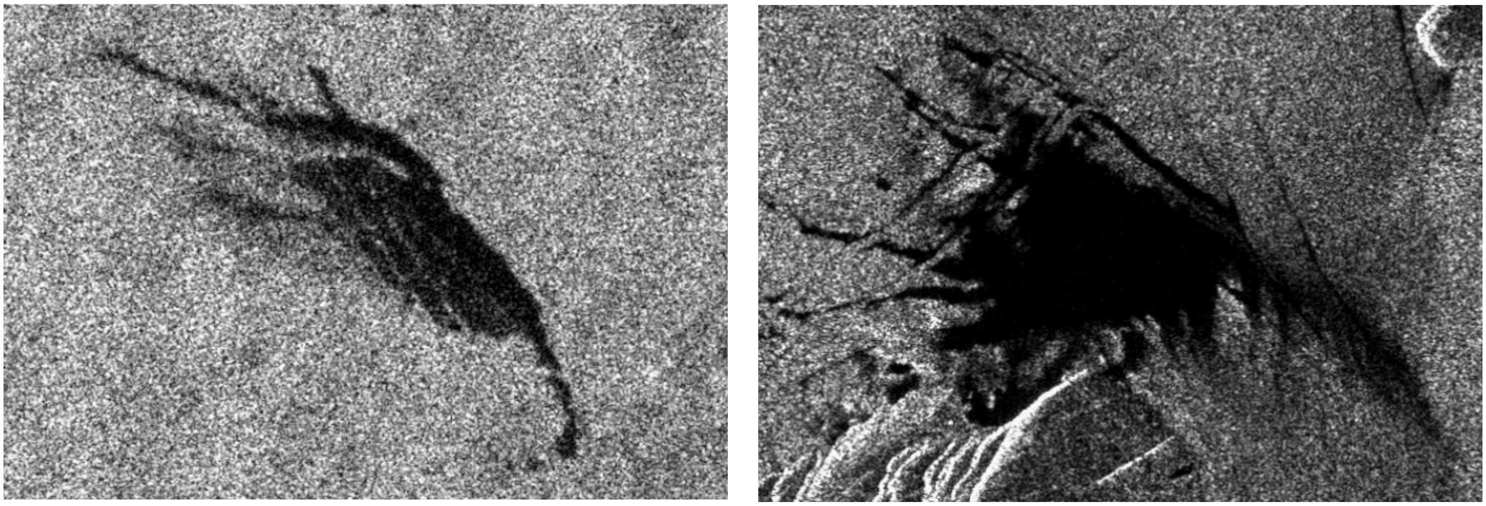

Figure 4. Comparison between a verified oil spill (left) and a verified look-alike film (right). Figure adapted from Stathakis *et al.* [11].

1.4 Dark Object Detection

Dark object detection finds possible oil spill candidates in SAR imagery. Oil spills and look-alike biogenic films are seen as dark spots on the ocean surface because they both dampen the radar's emitted signal. Since reflectance from dark objects is lessened, there is a lower backscatter intensity compared to the surrounding sea. Figures 1 and 4 show example dark objects of verified oil spills and verified look-alikes.

Conventional statistical approaches to dark object detection involve finding a local detection threshold value that separates dark object pixel values from everything else. Del Frate *et al.* [3] implemented object edge detection using histograms of the pixel data distribution. A recent study, described in Solberg [10] uses an adaptive threshold that automatically adjusts to be a certain distance from the mean backscatter value in a small viewing window. Kartathanassi *et al.* [5] uses a similar thresholding method that instead adapts to the overall brightness and contrast of SAR image.

This study uses a simplified thresholding algorithm involving two main steps. The first is a fuzzy C-means clustering technique that finds the main data cluster's mean and standard

#	Features	Oil Spills		Look-alikes	
		Mean (μ)	St. Dev (σ)	Mean (μ)	St. Dev (σ)
1	Area (A)	30471.09	55930.64	47242.99	116988.16
2	Perimeter (P)	2812.06	3823.01	4233.16	5038.28
3	Perimeter to Area (P/A)	0.21	0.14	0.25	0.19
4	Complexity (C)	4.33	2.40	6.02	2.56
5	Shape Factor I (SP1)	5.72	4.80	4.06	4.51
6	Shape Factor II (SP2)	0.86	0.19	0.72	0.26
7	Object mean intensity (OMe)	43.57	18.07	46.40	15.39
8	Object standard deviation (OSd)	27.92	7.19	30.32	6.51
9	Object std. dev to mean ratio (OSd / OMe)	0.72	0.29	0.72	0.26
10	Background mean intensity (BMe)	115.94	18.56	113.97	17.52
11	Background standard deviation (BSd)	44.28	8.52	46.80	7.81
12	Background std. dev to mean ratio (BSd / BMe)	0.39	0.08	0.42	0.07
13	Ratio of std. dev to mean ratios	1.84	0.62	1.73	0.54
14	Mean contrast (BMe – OMe)	72.36	20.67	67.58	21.27
15	Max contrast (ConMax)	113.08	18.02	112.83	17.56
16	Mean contrast ratio (OMe / BMe)	0.38	0.14	0.41	0.14
17	Std. dev contrast ratio (OSd / BSd)	0.63	0.10	0.65	0.10
18	Local area contrast ratio (ConLa)	0.39	0.13	0.45	0.14
19	Mean border gradient (GMe)	79.85	14.38	79.11	11.66
20	Border gradient std. dev (GSd)	226.62	34.11	234.92	18.72
21	Max border gradient (GMax)	37.90	7.87	38.24	7.50
22	Mean difference to neighbors (NDm)	35.86	12.44	31.84	13.36
23	Spectral texture (TSp)	33.84	8.44	36.37	7.58
24	Shape texture (TSh)	0.22	0.01	0.23	0.01
25	Mean Haralick texture (THm)	44.52	17.69	47.42	14.85

Table 1. Statistical parameters for features of oil spills and lookalikes, adapted from tables in Del Frate *et al.* [3] and Toupouzelis *et al.* [13].

deviation. This step's purpose is to prevent false detection caused by outliers. The second step calculates the threshold based on the statistics (mean, standard deviation) of the cluster. Every value below the threshold is labeled a "dark object" pixel, identified in Figure 2 by the values highlighted in red. The end result of dark object detection on the image in Figure 1 is shown in Figure 3. From there, the image is segmented into individual dark objects such as the one boxed in white.

1.5 Feature Extraction from Dark Objects

Features are numerical values that describe the physical and geometric qualities of detected dark objects. Geometric features include the area, perimeter and shape factors of the formations, and physical features include the mean backscatter values of the object and of the surrounding

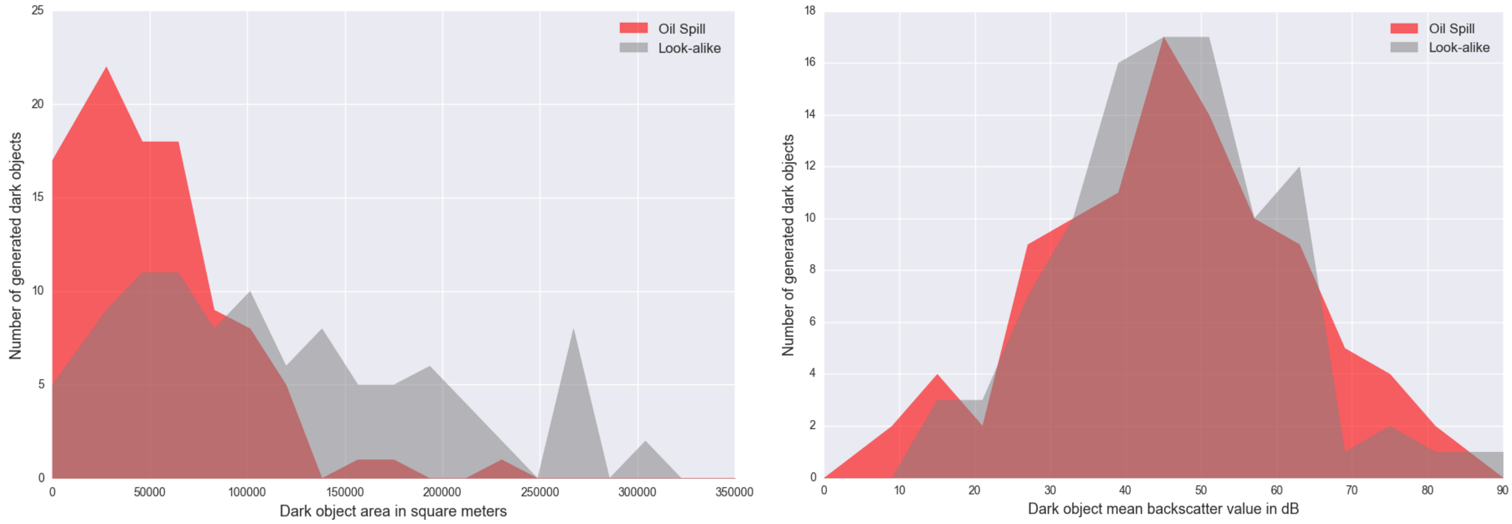

Figure 5. *(left)* Distribution for the areas of detected dark objects (in m^2), with oil spills red and look-alike objects in grey. *(right)* Distribution for the mean of object backscatter intensity values (in decibels) in dark objects.

ocean. These features are normally what manual oil spill inspectors would look for, and some or all of the features listed in Table 1 have been used in previous studies.

2 Methodology

Remote sensing applications for oil spill detection generally follow three steps: 1) dark object detection, 2) feature extraction from dark objects and 3) classifying the dark objects as either oil spills or lookalikes [12]. This study followed procedures from literature for steps one and two and focuses on improving the last step by using a genetic algorithm to optimize classification.

2.1 Dataset

Simulated data were generated from published statistics found in Toupouzelis *et al.* [13] and Del Frate *et al.* [3] due to the lack of verified oil spill datasets. A total of 400 labeled dark objects (200 oil spills and 200 lookalikes) were initialized as the training dataset, and a total of 200 labeled dark objects (100 oil spills and 100 lookalikes) were set aside as the testing dataset. For each spill and lookalike, the features were generated randomly along a Gaussian distribution, using the statistics in Table 1.

Figures 5 shows the distribution of the area (A) and mean of object backscatter intensities (OMe) for the generated dark objects in the testing dataset. The large overlap between the distributions shows that classifying between spills and lookalike objects is difficult (refer to Figure 4 for comparison). However, some features such as area have less overlap than others, meaning that those features will be more useful in determining which distribution the dark object falls under, whether the numerical inputs match the oil spill distribution or the look-alike distribution.

2.2 Classification

This study determines whether a detected dark spot is an oil spill or a lookalike film using a single layer artificial neural network (ANN). The ANN model implements a softmax regression that multiplies the input features by weight coefficients and outputs either a zero (lookalike) or a one (oil spill). Since this study does not focus on the classification algorithm itself, the simplest and least computationally expensive single-layer ANN is preferred over a multilayer perceptron network outlined in Del Frate *et al.* [3]. The ANN is trained for 2,000 iterations over the training dataset, and the accuracy is produced by classifying the testing dataset with the trained model. Overall classification accuracy is defined as the proportion of objects correctly categorized by the ANN in the entire testing dataset. Oil spill detection accuracy, or the *true positive rate*, and lookalike detection accuracy, or the *true negative rate*, are the proportion of oil spills and the proportion of look-alikes correctly classified in the testing dataset, respectively.

2.3 Feature Optimization using a Genetic Algorithm

Conventional research on oil spills classify between oil spills and lookalikes based on *arbitrarily selected features*. Some of these features hold little to no weight in classification and add another layer of noise, lowering the overall detection accuracy. In many cases, using every feature does not make sense when classification can be improved by only using a subset of features as inputs.

This study explores novel approaches to utilize a genetic algorithm in optimizing feature usage since the numerical features associated with a dark formation are ultimately the deciding factors for the classification algorithm. A genetic algorithm finds an optimal or near-optimal solution within a given search space by mimicking the natural processes of competition, reproduction, mutation etc.

This study uses a binary representation for the 25 features: a 0 means that the feature is excluded from the classification process, and a 1 means that the feature is included. A solution, or *chromosome*, is a set of 25 binary values, where each 1 represents a feature that will be passed into the classification ANN. Figure 6 shows the lifecycle of the genetic algorithm: the algorithm will continue until it has reached the specified maximum number of generations or when the algorithm has successfully converged on a chromosome.

2.3.1 Initialization and Calculating Fitness

The genetic algorithm in this study is initialized with a population of 20 randomly generated chromosomes. After training and testing the classification model with the randomly selected features, each chromosome is assigned a fitness value that is based on a predefined fitness function proposed by Siedlecki *et al.* [9]. Lower fitness values are associated with better solutions. The following function outputs the fitness of a chromosome x, where n represents the number of features used, r is the error rate (1 – overall accuracy) and m and t are experimentally defined constants that control the weight of the accuracy term. The genetic algorithm aims to minimize the fitness function, which is achieved in two ways: 1) a solution reduces the number of features used or 2) a solution increases the classification accuracy.

$$F(x) = n_x + \frac{e^{\frac{r-t}{m}} - 1}{e - 1}$$

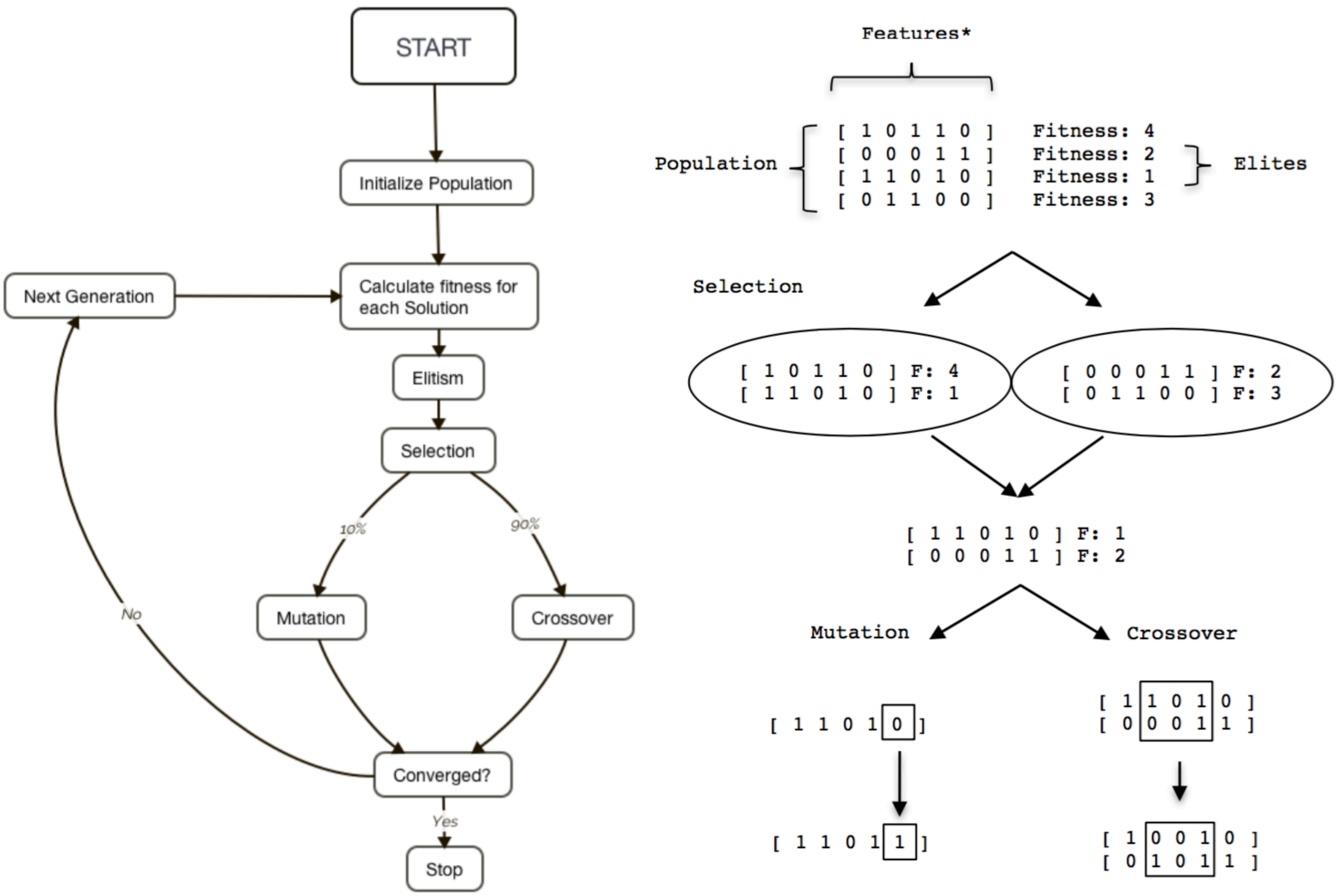

Figure 6. Genetic algorithm overview
(left) Lifecycle genetic algorithm operations: elitism, selection, crossover and mutation.

(right) Simplified examples for each process in the genetic algorithm:

1. Arbitrary fitness values were assigned to the chromosomes to show relative comparison (lower fitness values mean better solutions).
2. Selection chooses the better solution out of a tournament of size 2.
3. We see an index invert from a ‘0’ to a ‘1’ at a random index during *mutation*.
4. We see two chromosomes swap a randomly selected interval during *crossover*.

*There are 4 chromosomes and 5 features in the examples. The actual algorithm uses a population of 20 chromosomes with 25 features each.

The initial set of 20 chromosomes is referred to as the parent population of generation 0; each iteration of a genetic algorithm is called a *generation*. The 0^{th} generation's population will undergo a series of operations to produce child chromosomes, and the resulting set of 20 child chromosomes becomes the next generation's parent population. The following sections will describe the four main biological operations used to evolve a population of chromosomes: *elitism*, *selection*, *crossover* and *mutation*.

2.3.2 Elitism

Elitism ensures that the best solutions of each generation will make it into the next iteration. Without this process, the genetic information of good solutions may, by chance, be discarded from the gene pool. The progress made on those chromosomes would be lost, making it extremely difficult for the algorithm to eventually converge on an optimal solution. In this study, the top 10% (two) of solutions are considered the "elites," passed on to the following generation immediately.

2.3.3 Selection

Selection simulates competition in the mating process of reproduction. We used a tournament selection method to choose "male" and "female" partners. Each individual partner is selected by first creating a tournament consisting of 2 randomly chosen chromosomes. Then, the solution with the lower fitness is selected. The tournament size of 2 was determined after experimentation. In general, a larger tournament size exerts more competitive pressure on the population, which limits genetic diversity [1, 14]. Diversity can be lost in two ways: 1) a solution does not participate in any tournament due to random chance, and 2) a solution loses in the tournament that it is part of [7]. Thus, if each tournament consisted of 4 chromosomes instead of 2, a smaller number of unique chromosomes would contribute to diversity since it would be more likely for the same chromosomes to be selected and paired with each other.

2.3.4 Crossover and Mutation

Crossover mimics the biological process of meiosis (where the parent chromosomes will cross to form the offspring's chromosomes). The present study implements a two-point crossover technique that swaps the information of two parent chromosomes in a randomly defined interval, producing two child chromosomes. The crossover rate is 90%, meaning that a majority of male-female chromosome pairs will exchange genetic information in this fashion. The 10% of solutions that do not crossover are instead subject to single index mutation, which inverts the solutions at random indices (i.e., a "1" becomes "0" and vice versa).

We hypothesize that crossover plays the largest role in evolving new, robust solutions, since a hybrid of two existing solutions has the potential to improve old chromosomes. Crossover works hand in hand with selection since the population is already filtered to exclude relatively bad solutions. Mutation is usually responsible for removing unnecessary features. If the classification accuracy produced from a solution is the same regardless of the presence of a feature, then that feature is not needed as an input.

The processes of selection, crossover and mutation go on until 20 children are produced (including the "elites"), and those 20 chromosomes become the next generation's population. The process starts all over again by reassigning fitness values. Each trial of the algorithm runs for 50 generations.

2.3.5 The Local Maximum Problem

One problem that hinders a genetic algorithm's ability to optimize is *premature convergence*. This issue is closely linked to the loss of genetic diversity in the population, since new solutions cannot be easily evolved when all or most of the chromosomes are identical. When either selection pressure or elitism is too extreme, the genetic algorithm will quickly (in under 10 generations) converge to a local maximum. The solution that the algorithm converges to does not reflect the best set of features that can be used to classify between oil spills and lookalikes.

This study implements a novel technique called *adaptive chromosome replacement (ACR)*. When many individuals share the same genetic information, it is very likely that the identical chromosomes will be paired in crossover. This defeats the purpose of the crossover operation because the children produced will be exact replicas of the parents. ACR solves this issue by introducing new, randomly generated individuals whenever there is little deviation in the population. This technique was inspired by random offspring generation as described in Rocha *et al.* [8].

During the first half of a 50 generation trial, ACR uses a double replacement method when the standard deviation of classification accuracies is under a given threshold. Otherwise, it uses a single replacement method. Double replacement means that, before crossover, if the pair of chromosomes are identical, each individual in the pair will be replaced by a randomly generated solution. Single replacement means that in the same scenario, only one chromosome will be exchanged. In a real world scenario where a population of organisms becomes too similar in genotype, the population is highly susceptible to factors like disease. Thus, there is a need for rapid mutation and evolution in order for the population to survive: *ACR mimics this natural phenomenon by inserting random genes into a stagnating population.*

During the second half of the trial, ACR uses single replacement whenever double replacement would apply. When there is adequate deviation within the population, ACR swaps one of the chromosomes with another solution *in the population* instead of using single replacement. Through experimentation with ACR, it is clear that double replacement is effective at maintaining genetic diversity. However, it is extremely difficult to converge when good solutions are being discarded constantly, which is why single replacement is preferred.

3 Results

We ran the genetic algorithm 15 times and recorded the best solution from each trial in Table 2. For each solution, we trained the classifier using only the features marked with a "1." For example, the classifier for the first solution in Table 2 would only use the dark object's area, perimeter, perimeter to area ratio etc. After the classification model was trained using the training dataset of 200 oil spill objects and 200 lookalike objects, we used it to classify the testing dataset of 100 spills and 100 lookalikes. We then recorded the fitness of the solution as it pertains to the genetic algorithm, the overall accuracy, the true positive rate (oil spill accuracy) and the true negative rate (lookalike accuracy). The discrepancy between the true positive rate and the true negative rate is unusual (for example, 91% compared to 75% in trial 4). We believe that the true negative rate and overall accuracy would increase with a more complex neural network (with convolution and multiple hidden layers).

3.1 Summary and Comparisons

The best overall accuracy recorded was solution 7 in Table 2, where 84% (168 out of 200) of dark objects were correctly classified. This solution classified 87% (87 out of 100) of oil spills correctly and 81% (81 out of 100) of lookalikes correctly. The best true positive rate among the 15 trials was 93%, the best true negative rate was 88%. The *average of the oil spill detection accuracies* among the 15 solutions was 88.2%.

As a baseline for classification improvement, we trained and tested the classifier using all 25 features. This gave an oil spill detection accuracy of 80%, a lookalike detection accuracy of 79% and an overall accuracy of 79.5%. A two proportion z-test to compare the *oil spill detection* accuracies with and without the genetic algorithm showed that the improvement (from 80% to an average of 88.2%) was statistically significant ($p = 0.0559$) when $\alpha = 0.1$. In this case, comparing the true positive rate is most appropriate, since it is always better to misclassify a lookalike than to misclassify an oil spill.

We compared our results to other machine learning approaches that did not have a feature selection (genetic algorithm optimization) component. Del Frate *et al.* [3] implemented a multilayer perceptron network using 11 features to classify 139 dark formations. Their neural network classified 82% of oil spills and 90% of lookalikes correctly. Solberg [10] introduced a statistical model trained on a database of 7,051 dark objects. Their results showed an accuracy of 78% in oil spill classification (29 out of 37) and of 99% in lookalike classification (12,033 out of 12,110) [12]. Again, a one-tailed hypothesis test against the oil spill detection accuracy for both of these studies showed that our oil spill classification results are statistically significant improvements ($p = 0.0951$ against Del Frate *et al.* [3], $p = 0.06681$ against Solberg [10]).

3.2 Analysis of the Genetic Algorithm

Figure 7 shows a plot of the mean overall accuracy over 50 generations. The sharp increase in the first 5 generations shows that the genetic algorithm is able to quickly evolve the population using the four operations elitism, selection, crossover, and mutation as described before. The algorithm's progression slows down, gradually improving the mean accuracy until it reaches a near-optimal solution. The dips in mean accuracy show the effects of ACR. Every few generations, when the population of chromosomes becomes too similar, ACR will add back genetic diversity by introducing new, randomly generated solutions. Since it is very likely that the new solutions have lower overall accuracies, the mean accuracy as seen in the graph falls. After the initial dip in accuracy, the algorithm is able to evolve more accurate solutions with the diverse gene pool.

3.3 Feature Usage Ranking and Evolution

The feature usage for the 15 trial solutions is organized in Table 3. The most robust features are the ones used in multiple solutions, since the algorithm converges to those features regularly even though it is initialized in each trial with a completely random population.

Figure 8 is a more detailed look into the convergence process and the evolution of feature usage over the course of a single trial. In the beginning, solutions are completely random, so feature usage is spread out roughly evenly. Ten generations in, a few features colored in red become prominently used throughout the population. At thirty generations, the most robust features are used in the large proportion of solutions, although ACR still manages to maintain genetic diversity with features colored in blue. In the second half of the trial, the algorithm converges to a solution using 14 features.

4 Illustrations

	Feature																													
Trials	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Fitness	True Positive Rate (%)	True Negative Rate (%)	Overall Accuracy (%)	# Features
1	1	1	0	1	0	1	1	1	0	1	1	0	0	0	1	0	1	1	0	0	0	1	1	0	1	21.02	90	77	83.5	14
2	1	1	0	0	1	1	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	1	0	0	22.04	86	80	83	13
3	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	1	0	0	0	1	1	0	0	21.02	87	80	83.5	14
4	1	1	0	0	1	1	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	23.02	91	76	83.5	16
5	1	1	1	1	1	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	1	0	0	0	22.04	89	77	83	15
6	1	0	0	1	1	1	0	1	0	1	0	0	0	1	1	1	1	0	1	1	0	0	1	1	1	26.62	90	75	82.5	13
7	1	1	0	0	1	1	0	1	0	1	0	1	1	1	1	0	0	1	0	0	0	1	1	0	1	19.45	87	81	84	14
8	1	1	0	1	0	0	1	1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1	1	21.02	87	80	83.5	14
9	1	1	0	1	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	26.94	91	73	82	12
10	1	0	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	0	35.21	88	75	81.5	16
11	1	1	1	0	0	0	1	0	0	1	0	0	0	0	1	1	0	1	1	0	1	1	1	1	1	21.02	93	74	83.5	14
12	1	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	22.62	93	72	82.5	11
13	1	1	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	25.02	87	80	83.5	18
14	1	1	0	0	1	0	0	1	0	1	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	22.62	85	80	82.5	11
15	1	1	0	0	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	1	1	21.04	78	88	83	12

Table 2. The best solutions recorded from 15 different trials of the genetic algorithm. A highlighted '1' means that the feature at that index (see Table 1) is used in classification. 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%.

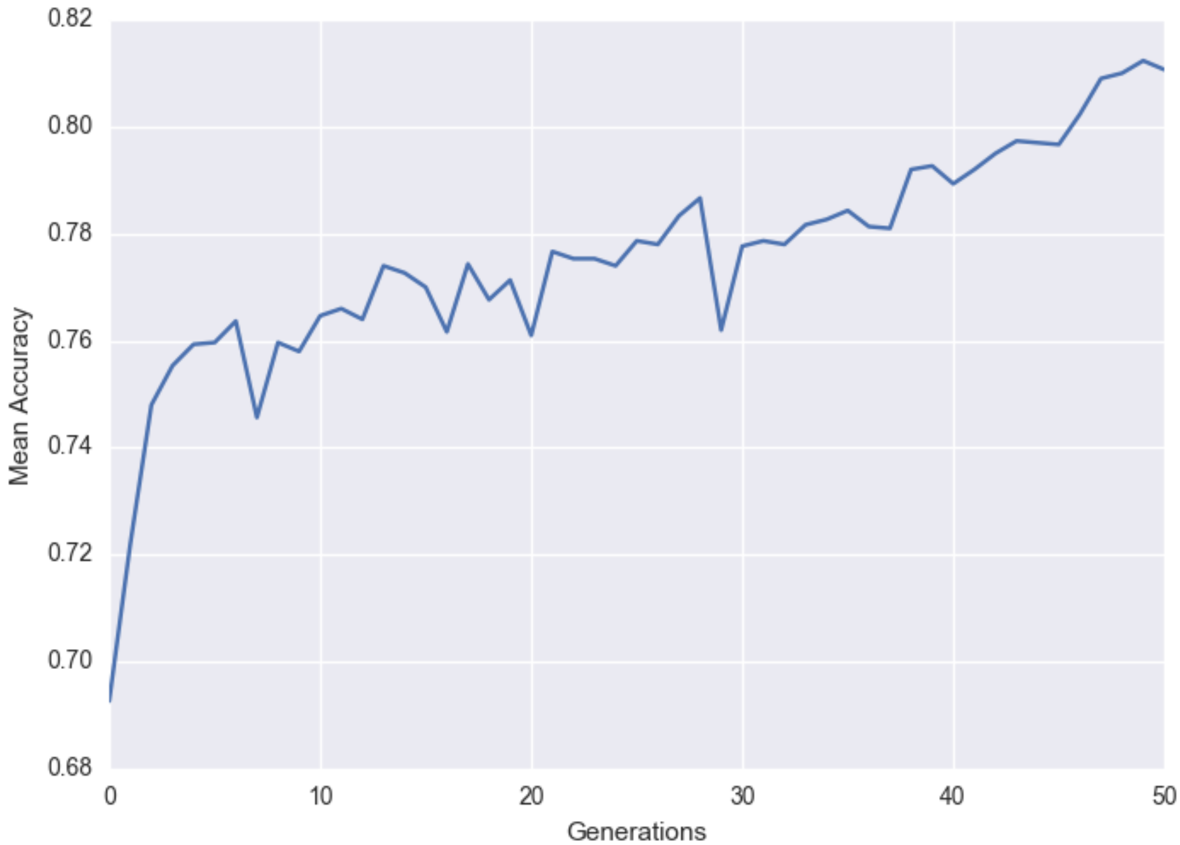

Figure 7. The mean classification accuracy of the population of solutions throughout the course of one trial (50 generations): ranges from around 69% to 81% in overall accuracy for an improvement of 12% accuracy from the original population.

Frequency Used	**Features**
81-100%	1, 10, 15, 22
61-80%	2, 23, 8, 5
41-60%	24, 25, 4, 6, 14, 17, 19
$\leq$ 40%	9, 11, 12, 13, 16, 18, 3, 7, 20, 21

Table 3. Feature ranking among the 15 trial solutions: the four best features, used in virtually every solution in Table 2 are area (1), background mean intensity (10), max contrast (15) and mean difference to neighbors (22). Refer to Table 1 for the feature descriptions that correspond to the numbers.

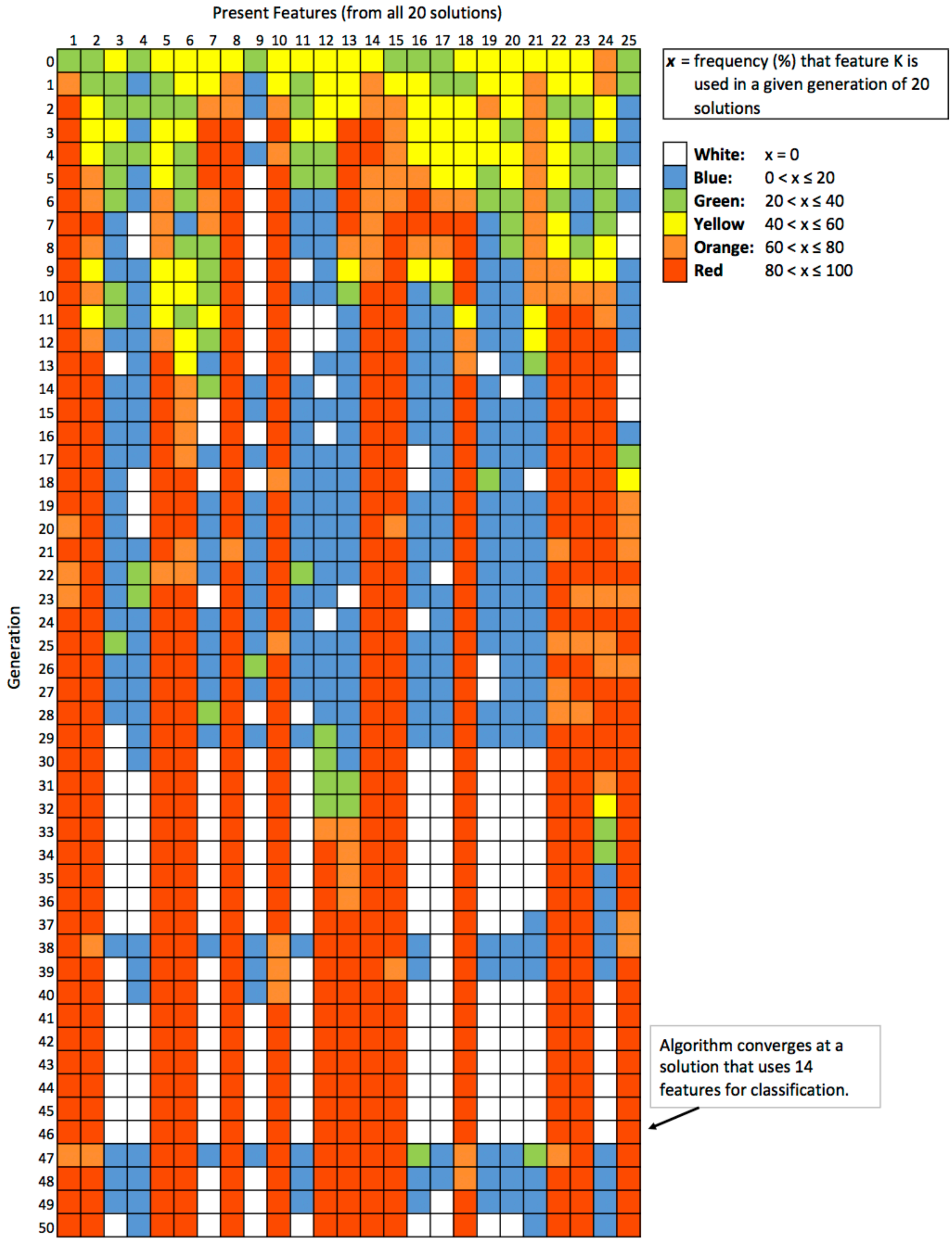

Figure 8. The evolution of feature usage over the course of a single run (50 generations) of the genetic algorithm. This particular run converged to solution 7 in Table 2. Features are labeled (1-25) according to Table 1. For every feature *K*, the frequency that it is used is calculated by taking the number of appearances in the population divided by the population size, and the corresponding cell is colored according to the legend on the right.

5 Conclusion and Future Work

Even though SAR is already used extensively in the field and has been recognized as an effective tool in assisting oil spill detection efforts, it isn't a perfect method. The accuracy of these methods can be further increased with new artificial intelligence and machine learning technologies. For example, it is possible to improve the classification algorithm by adding convolution and multiple hidden layers.

We would like to conduct more trials of the algorithm, using different methods for elimination of features. Taking the features and isolating the ones that provide the most success, as well as combining features that would have normally been eliminated may further improve the accuracy. Because the algorithm is still imperfect, data preprocessing methods may also be improved upon. Because the raw data does not include optical data, it is hard to discern between different types of oil spills. By introducing other features, such as wind patterns, it may become easier to create a distinction between oil spills and look-alikes.

Additionally, the data used in the tests were synthesized from previous studies and were not actual datasets from satellites. This may have led to, leading to different test results. The access to verified SAR data for oil spills is limited and, unfortunately, our application for a dataset hosted by the Canadian Space Agency was not approved. If satellite data were used in training and testing, our results would align with those of past results more accurately.

Overall, feature optimization using a genetic algorithm proved to be an effective way to improve oil spill detection. The *adaptive chromosome replacement* technique also counteracted the genetic algorithm's natural tendency to converge quickly and can be studied further in future research. The implementation of ACR raises another area for further research, which is the addition of other operations to the genetic algorithm to extend its functionality.

References

1. Blickle, T. and Thiele, L., “A Comparison of Selection Schemes used in Genetic Algorithms”, *TIK-Report*, Zurich, 1995

2. D’andrea, Mark A. and Kesava, G. R., “Crude Oil Spill Exposure and Human Health Risks”, *Journal of Occupational and Environmental Medicine*, 56.10 (2014), pp. 1029-041

3. Del Frate, F., Petrocchi, A., Lichtenegger, J. and Clabresi, G., “Neural networks for oil spill detection using ERS-SAR data.” *IEEE Transactions on Geoscience Remote Sensing,* 2000, 5, 2282-2287.

4. Jenssen, Bjørn Munro. “Review Article: Effects of Oil Pollution, Chemically Treated Oil, and Cleaning on Thermal Balance of Birds.” Environmental Pollution 86, no. 2, 1994, pp. 207-15.

5. Karathanassi, V., Topouzelis, K., Pavlakis, P. and Rokos, D., “An object-oriented methodology to detect oil spill.” *Int. J. Remote sens.* 2006, 27, pp. 5235-5251

6. Liang, S., Li, X.W. and Wang, J., “Advanced Remote Sensing”, Amsterdam: Academic Press, 2012. ISBN: 978-0-12-385954-9

7. Razali, N.M. and Geraghty, J., “Genetic Algorithm Performance with Different Selection Strategies in Solving TSP”, *Proceedings of the World Congress on Engineering*, Vol II WCE 2011, July 6 - 8, 2011, London, U.K.

8. Rocha, M. and Neves, J., “Preventing Premature Convergence to Local Optima in Genetic Algorithms via Random Offspring Generation”, *Proceeding IEA/AIE '99 Proceedings of the 12th international conference on Industrial and engineering applications of artificial intelligence and expert systems: multiple approaches to intelligent systems*, pp. 127-136

9. Siedlecki, W. and Sklansky, J., “A note on genetic algorithms for large-scale feature selection”, *Pattern Recognition Letters*, November 1989, vol. 10, Issue 5, pp. 335-347

10. Solberg, A., “Remote Sensing of Ocean Oil-Spill Pollution”, *Proceedings of the IEEE*, Oct. 2012, vol. 100, issue 10, pp. 2931-2945

11. Stathakis, D., Topouzelis, K., and Karathanassi, V., “Large-scale feature selection using evolved neural networks”, *Image and Signal Processing for Remote Sensing XII*, Proc. of SPIE, 2006, vol. 6365, pp. 636513-1 - 636513-9

12. Topouzelis, K., “Oil Spill Detection by SAR Images: Dark Formation Detection, Feature Extraction and Classification Algorithms”, *Sensors 2008*, vol. 8(10), pp. 6642–6659

13. Topouzelis, K., Stathakis, D. and Karathanassi, V., “Investigation of genetic algorithms contribution to feature selection for oil spill detection”, *International Journal of Remote Sensing*, 2009, vol. 30, issue 3, pp. 611-625

14. Whitley, D., “The genitor algorithm and selection pressure: Why rankbased allocation of reproductive trials is the best”, *In Proceeding of the 3rd International Conference on Genetic Algorithms*, 1989, pp. 116-123

15.http://www.nasa.gov/directorates/heo/scan/communications/outreach/funfacts/txt_passive_active.html

16. http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9567