Abstract—The complexity of existing crossover operators used in Genetic Algorithm is a critical factor that affects performance due to its negative impact on processing time. In this paper, a new crossover operator called Push and Pop Genes Exchange Operator (PPX) is introduced. The Improved Genetic Algorithm was applied and tested its accuracy for soil classification in Cagayan Valley. Result show an accuracy of 95% was successfully achieved in classifying a soil.

Index Terms—Push and Pop Genes Exchange Operator, Genetic Algorithm, Crossover operator, Confusion Matrix.

I. INTRODUCTION

Genetic Algorithm (GA) is a search-based optimization technique that uses the principles of genetics and natural selection to find optimal or near optimal solutions for difficult problems [1]. It is particularly useful for machine learning applications.

Genetic Algorithm (GA) allows a population composed of many individuals to evolve under specific selection rule to a state that maximizes the fitness [1]. Two essential operators in Genetic Algorithm (GA) are crossover and mutation. The crossover operator is responsible for producing offsprings by way of recombining information from two parents, thus providing major exploratory mechanism of the algorithm [1].

Mutation prevents convergence of the population by flipping a small number of randomly selected bits ensuring the continuous introduction of variation [1]. The unique cooperation between crossover and mutation, together with selection, provides the driving force behind GA [2].

The complexity of crossover operators play an important role in searching and providing solution to a problem since a more complex crossover operator results in longer processing times [10]. This becomes even more pronounced when dealing with search problems of greater complexities [6][7]. This paves way to the development of a new, less complex, and simplified crossover operator and named as Push and Pop Genes Exchange operator which follows the concept of Last In First Out (LIFO) approach.

This research work attempts to apply the Improved Genetic Algorithm using New Crossover Operator and it specifically aims to; (1) Simulate the Improved Genetic Algorithm for soil classification using a program interface and (2) Test the accuracy of the Improved Genetic Algorithm using a Confusion Matrix.

II. RELATED LITERATURE

This section presents the Improved Genetic Algorithm using a new crossover operator named Push and Pop Genes Exchange operator and a brief description of confusion matrix as a tool for evaluating the accuracy of the algorithm.

A. Push and Pop Genes Exchange Operator

A new PPX crossover operator was introduced into GA in order to speed up the crossover process and reduce processing time. PPX follows the concept of stacking using a Last In First Out (LIFO) approach. The proposed modified algorithm consists of five steps, as follows:

1. Create a stack
2. Repeat
   Push element into the stack
   If the stack is full
   End
3. Until loop=numElement
4. Repeat
   Pop the element from the stack
5. While stack<>empty

In the initial step, an n number of elements (or parents) for the crossover process are selected. In the succeeding steps, a collection of elements are subjected to two principal operations: push, and pop, which remove the most recently added element that has not yet been removed. The order in which the elements come off the stack is indicative of its LIFO approach.

The restrictions that apply to the stacks are shown in Figures 1 and 2 showing the elements before and after the crossover process, respectively. For elements A, B, C, D and E added to the stack in that order, element E is first to be removed since it was the last element inserted into the stack in consonance with LIFO.

Fig. 1. Elements before the crossover process.
The dataset used was a real soil testing data results taken from the Department of Agriculture in Cagayan Valley. A total of 50 tests were taken to serve as a basis for simulating the classification process use in this paper. Simulations were performed in the program interface. Training set for classification is pre-loaded which is composed of a total of 16 different soil series known to be as classes having different attributes. These classes will be converted into a series of binary string with a fixed length of 21 bits which would serve as a rule or pattern in classifying the soil. Selecting the attributes based on the data collected from the soil test is the first step. Once all set, the classification process can be started. The application will process the inputted values which would serve a parameter in soil classification process. And finally, a result will be displayed in the Result area of the application providing details about the best classification of the soil according to the patterns or rule found in the training set.

An actual output of the program interface is shown in Figure 4.

III. RESULTS AND DISCUSSIONS

Classifying the soil according to series was simulated in the developed program interface. The program interface was designed and developed using the C# programming language as shown in Figure 3.

The program composes of different functionalities to classify soil according to series based on the inputted value of attributes or properties of a soil like the color, texture, PH Level, fragments, and mottles.

1. **Feature Selection** – allows the user to select feature of the soil to be classified.
2. **Result** – provides a view of the classification result
3. **Crossover Result Viewer** – provides visual representation of crossover processing results
4. **Identify Soil Series Button** – performs the classification process

Testing the Improved Genetic Algorithm is very essential to determine its accuracy in terms of classifying soils in the region of Cagayan Valley.

Figure 5 shows the corresponding result of the classified soil result from 100 test result taken from the Department of Agriculture in Cagayan Valley. 2 soils were classified as Bantog instead of Alaminos, 1 was classified as Bago instead of San Manuel, 1 was classified as San Juan instead of Ilagan, and 1 was classified as Quinga instead of Tagulud. This misclassification factor was due to very close value of the different attribute of the soil.

As a final result, using 100 test results of soil taken from the Department of Agriculture in Cagayan Valley, the Improved Genetic Algorithm was able to determine the best soil classification for a soil having a color of black/grayish brown, texture of clay, PH level between 4.5 to 5.5, fragments of limestone, and a mottle of spots of red; manganese (bagiing) and iron.
Genetic Algorithm for classifying a soil had a favorable accuracy rate of 95% in determining the classification of the soil with only 5% of misclassifying them.

IV CONCLUSION

This paper presented the application of the Improved Genetic Algorithm for soil classification in Cagayan Valley. The simulation and testing was made possible using the developed program interface. Actual dataset was used in obtaining the results of this paper. The output of the application shows that the Improved Genetic Algorithm for soil classification provides a better and faster way of classifying a soil. A test was also conducted to determine the quality and accuracy of the Improved Genetic algorithm in terms of classifying a soil in Cagayan Valley.

REFERENCES