An Analysis of Integration of Hill Climbing in Crossover and Mutation operation for EEG Signal Classification

Arpit Bhardwaj
Indian Institute of Technology Indore
Indore,India
phd12110102@iiti.ac.in

Aruna Tiwari
Indian Institute of Technology Indore
Indore,India
artiwari@iiti.ac.in

M.Vishaal Varma
Indian Institute of Technology Indore
Indore,India
cs1100121@iiti.ac.in

M.Ramesh Krishna
Indian Institute of Technology Indore
Indore,India
ee1100220@iiti.ac.in

ABSTRACT

A common problem in the diagnosis of epilepsy is the volatile and unpredictable nature of the epileptic seizures. Hence, it is essential to develop Automatic seizure detection methods. Genetic programming (GP) has a potential for accurately predicting a seizure in an EEG signal. However, the destructive nature of crossover operator in GP decreases the accuracy of predicting the onset of a seizure. Designing constructive crossover and mutation operators (CCM) and integrating local hill climbing search technique with the GP have been put forward as solutions. In this paper, we proposed a hybrid crossover and mutation operator, which uses both the standard GP and CCM-GP, to choose high performing individuals in the least possible time. To demonstrate our approach, we tested it on a benchmark EEG signal dataset. We also compared and analyzed the proposed hybrid crossover and mutation operation with the other state of art GP methods in terms of accuracy and training time. Our method has shown remarkable classification results. These results affirm the potential use of our method for accurately predicting epileptic seizures in an EEG signal and hint on the possibility of building a real time automatic seizure detection system.

CCS Concepts

•**Computing methodologies → Genetic programming;** *Continuous space search;*

Keywords

Genetic Programming; Epilepsy; Crossover; Mutation; Fitness Function; Hill Climbing Search

GECCO '15, July 11 - 15, 2015, Madrid, Spain
© 2015 ACM. ISBN 978-1-4503-3472-3/15/07...$15.00
DOI: http://dx.doi.org/10.1145/2739480.2754710

1. INTRODUCTION

One of the most studied problem in machine learning is the Multi class classification [1]. It involves predicting the value of a categorical attribute based on the values of other attributes. Evolutionary algorithms (EA) are a class of computational techniques inspired by the Darwin's theory of natural evolution to solve the real life complex problems [2,3]. Genetic Programming (GP) [4] is an evolutionary learning methodology that offers a great potential for classification. GP is essentially considered to be a variant of genetic algorithms (GA) [4] that uses a complex representation language to codify individuals. GP is a very flexible heuristic technique that makes it very convenient to represent complex patterns in the form of trees and graphs, therefore working with various operations and functions become easier. Since GP is a search and optimization algorithm, it can be easily employed as a search algorithm for generating a classifier. The goal of GP is the evolution of computer programs [5]. A population of computer programs, which are feasible solutions to a given optimization problem are evolved by means of Darwinian principle of survival of fittest. A set of biologically inspired operations such as reproduction, crossover and mutation are used in evolving these computer programs. Reproduction [4], which replicates the principle of natural selection, selects and copies the best individuals to the next generation. In mutation, a node or a subtree of a chosen individual is supplanted by a randomly generated node or a subtree. Mutation is helpful in bringing diversity among individuals. Crossover combines two individuals to produce a new individual (offspring). The idea behind crossover is that the newly generated offspring might be better than both of the parents if it acquires required characteristics from each of the parent individuals. These operations are undergone in evolution by a user-defined probability for each operations.

The Crossover and Mutation operations could be sometimes devastating in nature [6], as the totally arbitrary choice in crossover and mutation couldn't guarantee the best choice, which ultimately results in decrease in classification accuracy. These operations are responsible for destroying good subprograms in evolved programs during the later stage of evolution. So, there is a need to find a better way to perform crossover and mutation operations in GP. Including

Hill Climbing search [7], a simple heuristic search method, with crossover and mutation operation could produce better individuals and improves classification accuracy. In Hill Climbing, a random solution is generated, and this process is iterated until a better solution is found. In this paper, we focus on integrating the Hill Climbing technique in crossover and mutation operator to the EEG signal classification, by proposing a novel Hybrid crossover and mutation operator, and analyze its improvements in terms of classification accuracy and computational time.

The most important application of EEG signal classification is the detection of Epileptic seizures in human brain [8]. The Human brain is the most complex and magnificent organ in the human body, and as a matter of fact, it is so complicated that it remains an exhilarating frontier. In short, the brain serves as the seat of human consciousness and dictates the behaviors that enables us to survive [9]. Electroencephalography [10] is the measurement of the continuous brain-wave patterns, or electrical activity of the brain and the chronicler of disorders of the nervous system. An EEG [11] is a recording of electrical signals from the brain with the placement of small metal discs called electrodes positioned on the human scalp. The changes and voltages in these electric signals are measured in terms of voltage fluctuations of brain. In recent years, EEG techniques are growing popular among researchers for the investigation of epileptic seizures in particular. These techniques may prove as quintessential tools when used in conjunction with more prevalent neuropsycological tests.

Epilepsy [12] is a critical neurological brain disorder originating from temporary abnormal discharges of the brain electrical activity, leading to uncontrollable movements and tremblings. It is a neurological condition in which an individual experiences prolonged abnormal bursts of electrical discharges in the brain and is characterized by unexpected recurrent seizures. A detailed analysis of the EEG records could provide a valuable insight in predicting seizures. Until now, the exact cause of epilepsy in individuals is unknown and the mechanisms that involved behind the seizures are little understood. Thus, efforts towards its diagnosis and treatment are of significant importance. Developing automatic seizure detection methods [13] is of great significance and can serve as first-rate clinical tools for the scrutiny of EEG data in a more unprejudiced and computationally coherent manner, since visual inspection for discriminating EEG signals is time consuming, imprecise and high costly, especially in the case of long-term recordings.

The classification of EEG signals into seizure and non-seizure using Genetic Programming is prominent in developing real time seizure detection systems, as we can attain high classification accuracies using GP. Integration of Hill Climbing in the crossover and mutation operation makes it feasible to achieve such high classification accuracies. In the previous paper [14], we proposed Constructive Crossover and Mutation Operator for Genetic Programming, which integrates the hill climbing with crossover and mutation operation. In this paper we propose a new Hybrid Crossover and Mutation operation and compare with the existing methods of GP in classifying the EEG signals.

The remainder of this paper is organized as follows: Section 2 overviews the essential background of the approach. It describes about the Empirical Mode Decomposition, Life Cycle of GP and Related work. Section 3 describes the proposed work, section 5 presents and analyses the experimental results and finally Section 6 draws conclusion and future work directions.

2. BACKGROUND

This section describes the essential background for this approach, which includes the Empirical Decomposition for Feature extraction, life cycle of Genetic Programming, the standard crossover and Mutation operation, the constructive crossover and mutation operation and also details the related work.

An epileptic seizure can be fore casted by classifying the incoming EEG signal into seizure and non-seizure. The below mentioned is the chronological approximation of the ongoing implementation done for building a real time Epileptic Seizures detection system.

- Extraction of Brain signal Input.
- Pre-Processing the extracted EEG signals using Empirical Mode Decomposition.
- Classification of EEG signals using Genetic Programming.
- Build a Real Time Epileptic Seizure Detection System.

2.1 Empirical Mode decomposition

The feature extraction process is carried out using Empirical Mode Decomposition. The method Empirical Mode Decomposition (EMD) was initially proposed by Huang et al [15] which instructs to decompose a non linearly stationary signal into a superposition of natural modes, each of which could be easily analyzed for their instantaneous frequencies and bandwidths. Bajaj et al [16] proposed the use of Empirical mode decomposition to extract the features to classify the EEG signal.

EMD is a method of breaking down a signal without leaving the time domain introduced for analysis of nonlinear and non- stationary signals. EMD decomposes any given data into intrinsic mode functions (IMF) [15] that are not set analytically and are rather determined by an analyzed sequence alone, each successive IMF contains lower frequency oscillations than the preceding one. On completion of this process, it would generate a set of bandwidth parameters which are attributed as features for the Genetic Programming (GP) classifier. A typical implementation of EMD for decomposition of an EEG signal involves the following steps:

- Calculation of IMF for each iteration using EMD on EEG signals.
- Apply Hilbert transform on the calculated IMF's during each iteration.
- Generation of a Bandwidth parameters viz. Frequency parameter and the Amplitude parameter as the features for the classifier from the Hilbert Transform.

These Bandwidth parameters constitute the input parameters for the Genetic Programming classifier. During the feature extraction, we select five IMF's from each EEG signal and remove the residue. In sum, a total of 10 features from each EEG signal (two from each IMF's) are extracted and chosen as input to the GP classifier.

2.2 Life Cycle of Genetic Programming

GP evolves a population of computer programs, which are feasible solutions to a given optimization problem, using the Darwinian principle of survival of the fittest. It works on a principle, for instance, if there exists several entities in a nature, the fitter entities survive and evolve at a higher rate; less fitter individuals survive, if at all at a lower rate. It is an extension of the Genetic Algorithms and was substantiated, endorsed and developed into a practical tool by John Koza [4] amongst a whole range of possible evolutionary algorithms. GP is a very flexible heuristic technique that makes it very convenient to represent complex patterns in the form of trees and graphs, therefore working with various operations and functions becomes easier. Each individual in the population is represented as a tree with functions and terminals pertinent to a given problem. The evolutionary GP life cycle is detailed as follows:

1. Generate the initial random population.
2. Evaluate the fitness measure of each individuals using training data.
3. Evolve the individuals to get a new generation using reproduction, crossover and mutation.
4. Terminate the GP process on finding a best individual, or else go to step 2 and repeat the process

2.3 Crossover and Mutation Operation

2.3.1 Standard Crossover

In standard crossover operation [17], two parent individuals are selected from the population for the crossover operation. A crossover point is then randomly selected in each of the two parents, then the subtrees below the crossover points are exchanged and two new child individuals are spawned. The two new generated children are then transferred to the next generation.

2.3.2 Standard Mutation

In standard Mutation, an individual is selected from the population and a node or subtree is randomly chosen and replaced with a randomly generated node or subtree.

2.3.3 Constructive Crossover

In constructive crossover [14], a local hill climbing is integrated with the crossover operation. Here individuals are split into N_r and N_c for reproduction and crossover respectively. Initially, the N_r individuals undergo reproduction. After the reproduction, the remaining N_c individuals undergo constructive crossover, where individuals are made into pairs and are subjected to perform the crossover operation. The generated offspring from selected couples are compared with the parents in terms of fitness. While comparing, if it is known that the offspring are better than their parents in terms of fitness, they are accepted. Otherwise they are rejected and the process is repeated till we get two individuals better than parent. In this manner, the local hill climbing method is integrated with the crossover operation and is repeated for all N_c individuals and generated offspring are placed in Crossover Offspring Table (COT). A COT is a table in which we store the offspring generated from the crossover operations in such a way that the offspring having the highest fitness remain in the first row, the second highest fitness offspring in second row and so on. The top N_cr offspring in terms of fitness, the globally prime offspring, are present on the top N_cr positions in the Crossover Offspring Table (COT) and the remaining lower level offspring in COT are deleted. The remaining N_m individuals are selected for mutation operation.

2.3.4 Constructive Mutation

In the constructive mutation [14], the individuals, left after reproduction and crossover operations are chosen to generate better offspring while compared to their parents by applying hill climbing search. We randomly choose a subtree in a parent and replace it with a newly generated subtree and repeat this process till we get better offspring than parent.The benefit of employing this constructive mutation technique is that the destructive nature of mutation operation is reduced by transferring the better individuals than parents to next generation and also the constructive mutation provide a wider exploration of the search space, help in not sticking in local optima. The individuals, who are not good at producing better offspring than parents during crossover are altered in this constructive subtree mutation operation.

The aim of the constructive crossover and mutation operation is to find children with better fitness and improve the classification accuracy.

2.4 Related Work

Till now, a numerous methods have been proposed to classify an EEG signal into seizure and non-seizures in order to prognosticate epileptic seizures. Panda et al [18] produced different features, such as energy, entropy and standard deviation from wavelet transformation to classify epileptic EEG signals using support vector machine (SVM). Liang et al [19] used combination of complexity analysis and spectrum analysis and entropy features. Ocak [20] proposed fourth-level wavelet packet decomposition for several frequency bands to differentiate normal and epileptic EEG signals. A wide range of feature extraction methods are proposed which are used in conjunction with a very renowned classification technique Artificial neural network (ANN), to detect the epileptic seizures [21–25]. More recently, a new technique for classification was proposed which used Empirical Mode Decomposition and LS-SVM to differentiate between seizure and non-seizure EEG signals [16]. Here, the features are extracted by decomposing the EEG signal into a set of IMF's and then are fed into LS-SVM classifier. However, Recently, evolutionary algorithms have been emerged as a promising technique for classification of medical data [26]. Genetic Programming, developed by John Koza [4] could be used for the EEG signal classification. Castelli et al. [27] introduced a new GP system that uses the concept of semantics to improve search effectiveness. Arpit et al. [14] presented the concept of Constructive Crossover and Mutation, where a local hill climbing search was integrated with the crossover and mutation operation to improve classification accuracy. However, this method does not guarantee the best solution.

The goal of this paper is to propose a hybrid crossover and mutation operator, which finds the best children generated from the parent and reduce the computational time to reach the desired accuracy. In this paper, we compare and analyze the proposed hybrid crossover and mutation operation, which uses both standard crossover and local hill climbing crossover operation.

3. PROPOSED WORK

Here, we propose to investigate the variations in the standard GP process and Constructive Crossover and Mutation GP [14] process and its applications in the epileptic seizure detection. In the proposed Hybrid Crossover and Mutatation Genetic Programming (HCM-GP) method, we eliminate the randomness of crossover operation and reduce the computational time by introducing a novel hybrid crossover operator. We bring in diversity among individuals by introducing hybrid mutation operator. The flow chart of the proposed HCM-GP method is shown in the Fig 1

3.1 Hybrid Crossover Operation

The proposed Hybrid Crossover could be regarded as the blend of standard crossover and the constructive crossover operation. In the constructive crossover, a local hill climbing technique is integrated with the crossover to improve the classifier. But introduction of constructive crossover leads to decrease in the computational speed, as a great deal of time is spent in selection of better offspring. In order to significantly reduce the computational time, we introduced the hybrid crossover operation.

In our proposed method, we select the best fitness N_r individuals from the total population for *reproduction*, the next N_c individuals for the crossover and the remaining N_m individuals for the mutation. Among the N_c individuals for the crossover, individuals are randomly chosen between standard crossover and local hill climbing crossover. The split up of individuals into 50-50 between hill climbing and standard crossover has shown promising results while experimenting the proposed crossover operation. Individuals selected for the local hill climbing crossover are made into pairs and are subjected to perform the crossover operation to generate offspring. Here, the generated offspring is calculated for its fitness measure and is compared with its parents. If the parents have higher fitness value than the offspring, crossover is again repeated on the offspring. Else the offspring are transfered to the next generation. Here, we limit the number of crossover iterations to 20 [28] in order to avoid infinite loops, particularly when we are unable to generate a better offspring. The local hill climbing crossover is similar to the constructive crossover. The remaining individuals perform the standard crossover operation and generate their offspring to maintain the diversity. Hence, we ensure that the classifiers with higher fitness are transferred to the next generation, while maintaining the diversity among the population. The proposed hybrid crossover, which is categorized as a blend of hill climbing and standard crossover is better than the standard crossover, because it applies global elitism to some extent, by selecting the best individuals from a set of combinations, while simultaneously bringing the diversity among those individuals. The hybrid crossover is better than the constructive crossover because it does not take large amounts of computational time unlike constructive crossover.

3.2 Hybrid Mutation Operation

In the proposed hybrid mutation operator, we apply a hill climbing search on the N_m individuals, left after reproduction and crossover operations to generate better offspring while compared to their parents. In the constructive mutation operation, the low performing individuals in the crossover are brought to the mutation, where as in the hybrid mutation, we assign a fixed number of N_m individuals to the mutation operation after the fitness evaluation ie., prior to the crossover operation. Here, a random subtree was replaced with a newly generated random subtree. This process is repeated till we reach an individual with better fitness value. The benefit of employing this hybrid mutation technique is that the destructive nature of mutation operation is eliminated by transferring the better individuals to next generation. Moreover, This hybrid mutation technique doesn't make the searching of the solution to stuck in local optima. And the other advantage is about the selection, where the parents of mutation are chosen subsequent to the crossover and reproduction operation, as a result providing a scope for the low fitness offspring to survive in the evolutionary world.

Algorithm 1 Algorithm for Hybrid Crossover

```
1: Begin
2: Generate initial classifier population (k).
3: Randomly select fixed percentages of the initial population for crossover crossover (P_c), mutation (P_m) and reproduction (P_r).
4: Select N_c individuals, based on fitness measure, after the reproduction.
5: Split the N_c individuals into two groups randomly.
6: On one half, apply Hill Climbing
7: for all Selected half of the Individuals do
8:     Take the parent pair and generate two better offspring from them by applying hill climbing search.
9:     Place the top offspring (includes the parents if they have a superior fitness after 20 attempts).
10: end for
11: On the other half, apply Standard crossover
12: for all Selected half of the Individuals for standard crossover do
13:     Take the parent pair and generate offspring from using standard crossover.
14: end for
15: The resulting N_c offspring will advance to the next generation.
```

Algorithm 2 Algorithm for Hybrid Mutation

```
1: Generate initial classifier population (k).
2: Take the individual (N_m) which are left after reproduction and crossover operations.
3: for all mutation individual (N_m) do
4:     Apply Hill Climbing method, that is take the parent and generate child from them till we get better child than parent by replacing a subtree with a newly generated subtree.
5: end for
6: Transfer the individuals to the next generation.
```

All in all, the introduction of these variations in the GP life cycle is essential for the EEG signal classification, as it greatly reduces the computational time. Hence, the proposed hybrid crossover and mutation can be employed as an auxiliary tool for the neurologists to determine onsets

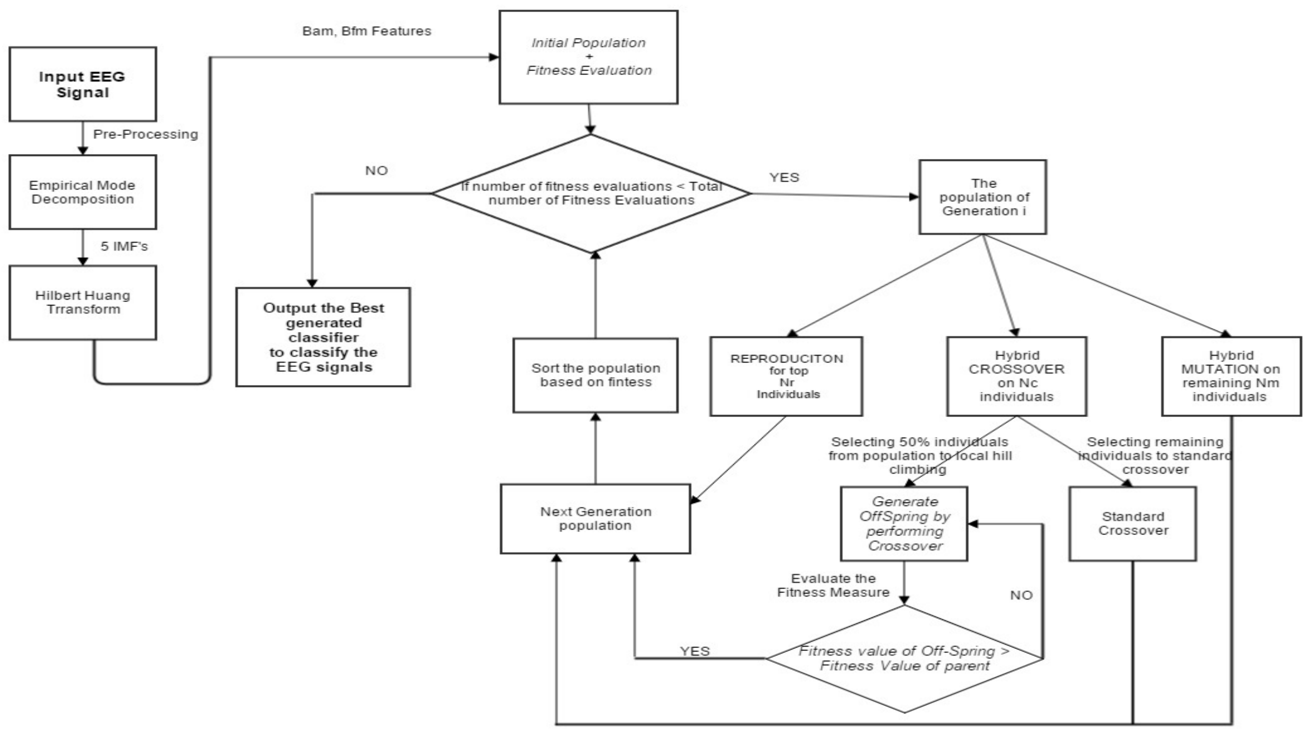

Figure 1: Flow Chart of the proposed Hybrid Crossover and Mutation GP

of seizures in analyzing 24 hr EEG recordings. In the next section, we demonstrate the results of the hybrid crossover and mutation operation and analyze the integration of hill climbing in the existing and proposed methods.

4. RESULTS AND DISCUSSIONS

This section presents the experimental results of the proposed Hybrid Crossover and Mutation. We analyze the effect of integration of local hill climbing in crossover and mutation by comparing the HCM-GP method with the existing ST-GP and CCM-GP methods.

4.1 Dataset Description

An EEG dataset, which is available on-line [29] was used for training, testing and evaluation of our method. In this dataset, the signals were recorded with a *128-channel* amplifier system as an average common reference. The analog data were digitized at 173.61 samples per second by a 12 bit A/D resolution with band-pass filter settings of 0.53-40 Hz (12 dB/oct). The dataset comprises of five different sets (denoted Z, O,N, F and S), each containing 100 signal channel EEG segments of 23.6 sec. duration. These signals were carefully chosen and cut out from continuous multi-channel EEG recordings after removing artifacts caused due to eye and muscle moments and power line interference. EEG signals in sets O and Z have been recorded from healthy volunteers through external surface electrodes using the standardized 10-20 electrode placement scheme. The volunteers were relaxed in an awake state with eyes open (set Z) and closed (set O). Sets F and N are obtained in seizure free intervals. Set F acquired from epileptogenic zone of the brain shows focal interictal activity; set N extracted from hippocampal formation of the opposite hemisphere of the brain hints the non-focal interictal activity, and set S were obtained from within the epileptic zone of a seizure activity. On an overall, Sets N, F and S originated from an EEG register of presurgical diagnosis. Sets Z and O are portrayed as Normal, sets N and F as interictal and Set S corresponds to an ictal State (Seizure). Here, The sets Z, O, N and F are combined and called as **Non-Seizure**, while the set S as **Seizure**. The goal of this experiment is to classify between seizure and non-seizure.

4.2 Performance Measures

The existing and proposed methods are evaluated by computing statistical parameters sensitivity, specificity and accuracy. Moreover, we also used Confusion matrix and ROC Curve for comparing the methods.The definitions of these performance measures are as follows:

1. **Accuracy:** Accuracy is the measure of the ability of the classifier to accurately classify the patterns.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (1)$$

2. **Sensitivity:** Sensitivity is the fraction of correctly detected positive patterns to the total number of actual positive patterns.

Table 1: Confusion matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \qquad (2)$$

3. **Specificity:** Specificity is the fraction of correctly detected negative patterns to the total number of actual negative patterns.

$$Specificity = \frac{TN}{TN + FP} \times 100 \qquad (3)$$

4. **Time required**: It is the time required to find the best individual.

5. **Confusion matrix**: Confusion matrix consists of information about actual and predicted classifications performed by a classifier. A confusion matrix used for our classifier, generally for two class problems is shown in Table 1.

6. **ROC curve** [30] is a fundamental diagnostic test evaluation and an excellent way to compare the performance of diagnostic tests. It is a two dimensional measure for the performance of classification and gives a complete sensitivity/specificity report. It also play an important role in comparing our test to the perfect test. ROC curve is plotted with *1-specificity* on the x-axis to *Sensitivity* on the y-axis.

4.3 Experimental Results

An important aspect underlying the HCM-GP method is that, it could be successfully modified to automatically detect highly suspicious seizure activities in brain through real time detection. In order to evaluate the performance of the methods, the dataset was divided into training and testing in several combinations namely 50-50, 60-40, 70-30 and 10-fold cross validation. Table 2 shows the partition into training and testing dataset for the classification into non-seizure and seizure (ZONF-S).

Table 2: Training and Testing dataset for the classification of EEG signals into Seizure and Non Seizure (ZONF-S)

Training-Testing Partition (%)	No. of samples in the set: Training Samples	ZONF	S	Testing Samples	ZONF	S
50-50	250	200	50	250	200	50
60-40	300	240	60	200	160	40
70-30	350	280	70	150	120	30
10-fold Cross Validation	450	360	90	50	40	10

Since, we are interested in investigating the integration of local hill climbing in GP life cycle, we analyzed the HCM-GP with Standard GP [4] and CCM-GP method [14], in terms of statistical parameters. The confusion matrix for the methods using various training-testing partitions are given in table 3.

Table 3: Confusion matrices for classifiers used for classification of EEG signals into Seizure and Non Seizure for different training-testing data

Classifier		50-50 ZONF	50-50 S	60-40 ZONF	60-40 S	70-30 ZONF	70-30 S	10 fold cross validation ZONF	10 fold cross validation S
ST-GP	ZONF	188	12	148	12	114	6	38	2
	S	8	32	3	37	3	27	2	8
CCM-GP	ZONF	196	4	156	4	118	2	39	1
	S	4	46	3	37	3	27	1	9
HCM-GP	ZONF	198	2	158	2	119	1	39	1
	S	1	49	1	39	1	29	0	10

Validation Techniques (Testing data)

Table 4: Comparison of HGP method with ST-GP and CCM-GP in terms of Performance measures

Method	Validation Technique	Sensitivity Min	Sensitivity Max	Specificity Min	Specificity Max	Accuracy Min	Accuracy Max	Time Required Min	Time Required Max
ST-GP	50-50	93.5	94.5	83.45	84.45	91.5	93	7.6	8.21
	60-40	92	94.5	92	94	92	94.5	8.09	9.10
	70-30	93.65	95.85	81	83	91.5	94.5	8.56	9.95
	10 fold cross validation	94	95.5	80	82	92	94.5	9.42	10.45
CCM-GP	50-50	95.5	98.65	90	93	95.5	97.5	5.50	6.26
	60-40	97	98.4	91.6	93.5	96	97.5	6.1	6.45
	70-30	97.85	99.5	89.5	92	96.65	97.75	6.95	7.45
	10 fold cross validation	97	98.5	90.5	92.5	96	97.5	7.65	8.13
HCM-GP	50-50	98.5	99.5	98.6	99.35	98.85	99.5	2.35	2.59
	60-40	98.875	99.45	98.5	98.85	98.75	99.65	3.12	3.48
	70-30	99.10	99.75	98.24	99.05	98.65	99.68	3.30	3.59
	10 fold cross validation	98.85	99.25	99.5	100	99.5	100	4.10	5.10

The sensitivity, specificity and accuracy and training time of the methods are quantified, noted and are shown in Table 4.

We also plotted the ROC curve using the above results. Fig 2 shows the ROC curve and area under the ROC curve for 50-50 training-testing data.

The above results demonstrate that our proposed method has achieved better classification accuracy than the other methods. It is also clear from the result, that our method requires lesser amount of time to reach the desired accuracy.

The high classification accuracy of our HCM-GP method compared to the other GP methods and consequentially we can affirm that this outweighs the standard GP and CCM-GP methods and we suggest that it can be used as a diagnostic decision support mechanism in the treatment of epilepsy patients. In our method, We integrated the local hill climbing in our crossover and mutation operation, which removed the randomness of the crossover operation and also guarantees a better solution, because it choses a better offspring than the parents, which inevitably increases the classification accuracy.

It is found that our method outperforms the Standard GP and CCM-GP methods. Our method improves the overall accuracy and speed by a fair amount. The standard GP can be destructive in nature. In standard crossover and mutation, rather than generating a offspring with better fitness, it could produce offspring with lesser fitness and transfer them to next generation. The CCM-GP acheived better accuracy than the Standard GP, its performance is low while com-

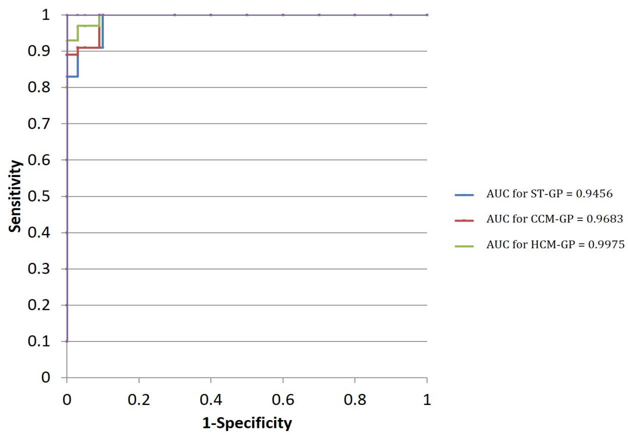

Figure 2: ROC Curve for 50-50 training-testing data

pared to HCM-GP. In CCM-GP, as all the individuals left after the reproduction undergo local hill climbing during the crossover, it takes a great deal of time to reach the desired accuracy. As, a result the CCM-GP method is time consuming, which is not ideal for real time detection of seizure. The proposed Hybrid Crossover searches for the best solution, in least possible time by merging both the standard crossover and constructive crossover. The Hybrid Mutation is responsible for improving the low performing individuals by bringing diversity among them.

Hence we can conclude that, though the integration of local hill climbing provides better solution, it is often time consuming, which is a major drawback for the CCM-GP method. Hence, we should be wise while integrating the hill climbing technique with the crossover and mutation. The proposed HCM-GP method, which uses local hill climbing on half of the individuals, still consumes a bit longer time than the Standard GP. But this is small price to pay, if it can significantly improve the classification performance. This makes the proposed HCM-GP framework a suitable tool to assist the experts by facilitating analysis of a patient's information and decreasing the time and effort required to accurately diagnose their patients.

5. CONCLUSION

An effectual real time detection of epileptic seizures is of paramount importance in clinical diagnosis of epilepsy. Genetic Programing has a huge potential in classifying the EEG signals to detect these epileptic seizures. Integration of local hill climbing technique with the crossover and mutation operations removes the destructive nature of crossover operation and improves the GP life cycle. However, we have to overcome a few limitations of the hill climbing search in order to build an effectual real time seizure detection system. The goal of this paper was to analyze the integration of local hill climbing method with crossover and mutation and to help overcome its limitations. The goal was successfully achieved by proposing hybrid Crossover and Mutation (HCM) operation. The purpose of the hybrid crossover and mutation is to attain the desired accuracy in least possible time. This approach uses the both Standard GP and CCM-GP. In the Hybrid Crossover, the first half of the individuals undergo conventional crossover operation, while the other half of the individuals undergo constructive crossover operation, where individuals which are only better than their parents are transfered to the next generation. The individuals which are left after the hybrid crossover and reproduction undergo hybrid mutation operation, where again a local hill climbing search is integrated with the mutation.

This approach was examined and compared with Standard GP and CCM-GP on a benchmark EEG signal dataset. The results suggest that the new approach outperformed the standard GP, while spending a bit longer time that the Standard GP. The new proposed operator achieves better classification accuracy in lesser time while compared to the CCM-GP. We can infer that the proposed HCM-GP method is considered as an efficient way to classify EEG signals and an inspiration for further developments in this area.

5.1 Future Work

Although this method improves the crossover and mutation operation and subsequently the GP life cycle, but still it does not provide a feasible solution.

The goal of this paper is to demonstrate the proposed HCM-GP and to analyze the integration of hill climbing with crossover and mutation with respect to EEG signal classification.So, the proposed method was only examined on EEG signal dataset and was not examined on other wider application tasks. We will further investigate this new classification approach on more general and wider range of pattern recognition problems such as Alzheimer's and Parkinson's diseases, breast cancer and diabetes detection and diagnosis in the future.

6. REFERENCES

[1] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Dept. of Informatics, Aristotle University of Thessaloniki, Greece*, 2006.

[2] Z. Michalewicz, *Genetic algorithms+ data structures= evolution programs.* Springer Science & Business Media, 1996.

[3] K. C. Tan, T. H. Lee, and E. F. Khor, "Evolutionary algorithms for multi-objective optimization: performance assessments and comparisons," *Artificial intelligence review*, vol. 17, no. 4, pp. 251–290, 2002.

[4] J. R. Koza, *Genetic Programming: vol. 1, On the programming of computers by means of natural selection.* MIT press, 1992, vol. 1.

[5] ——, "Genetic evolution and co-evolution of computer programs," *Artificial life II*, vol. 10, pp. 603–629, 1991.

[6] K. Yong11, "Improving crossover and mutation for adaptive genetic algorithm," *Computer Engineering and Applications*, vol. 12, p. 027, 2006.

[7] U.-M. O'Reilly and F. Oppacher, "Program search with a hierarchical variable length representation: Genetic programming, simulated annealing and hill climbing," in *Parallel Problem Solving from NatureâĂŤPPSN III.* Springer, 1994, pp. 397–406.

[8] M. During and D. Spencer, "Extracellular hippocampal glutamate and spontaneous seizure in the conscious human brain," *The lancet*, vol. 341, no. 8861, pp. 1607–1610, 1993.

[9] M. D. Bownds and D. Bownas, *The biology of mind: Origins and structures of mind, brain, and consciousness.* Fitzgerald Science Press Bethesda, MD, 1999.

[10] S. Sanei and J. A. Chambers, *EEG signal processing.* John Wiley & Sons, 2008.

[11] M. Teplan, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[12] P. Gómez-Gil, E. Juárez-Guerra, V. Alarcón-Aquino, M. Ramírez-Cortés, and J. Rangel-Magdaleno, "Identification of epilepsy seizures using multi-resolution analysis and artificial neural networks," in *Recent Advances on Hybrid Approaches for Designing Intelligent Systems.* Springer, 2014, pp. 337–351.

[13] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[14] A. Bhardwaj, A. Tiwari, M. V. Varma, and M. R. Krishna, "Classification of eeg signals using a novel genetic programming approach," in *Proceedings of the 2014 conference companion on Genetic and evolutionary computation companion.* ACM, 2014, pp. 1297–1304.

[15] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[16] V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure eeg signals using empirical mode decomposition," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 6, pp. 1135–1142, 2012.

[17] R. Poli and W. B. Langdon, "On the search properties of different crossover operators in genetic programming," *Genetic Programming*, pp. 293–301, 1998.

[18] R. Panda, P. Khobragade, P. Jambhule, S. Jengthe, P. Pal, and T. Gandhi, "Classification of eeg signal using wavelet transform and support vector machine for epileptic seizure diction," in *Systems in Medicine and Biology (ICSMB), 2010 International Conference on.* IEEE, 2010, pp. 405–408.

[19] S.-F. Liang, H.-C. Wang, and W.-L. Chang, "Combination of eeg complexity and spectral analysis for epilepsy diagnosis and seizure detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 62, 2010.

[20] H. Ocak, "Optimal classification of epileptic seizures in eeg using wavelet analysis and genetic algorithm," *Signal processing*, vol. 88, no. 7, pp. 1858–1867, 2008.

[21] I. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of eeg signals using wavelet coefficients," *Journal of neuroscience methods*, vol. 148, no. 2, pp. 113–121, 2005.

[22] N. F. Güler, E. D. Übeyli, and İ. Güler, "Recurrent neural networks employing lyapunov exponents for eeg signals classification," *Expert Systems with Applications*, vol. 29, no. 3, pp. 506–514, 2005.

[23] K. Aslan, H. Bozdemir, C. Şahin, S. N. Oğulata, and R. Erol, "A radial basis function neural network model for classification of epilepsy using eeg signals," *Journal of medical systems*, vol. 32, no. 5, pp. 403–408, 2008.

[24] L. Guo, D. Rivero, J. Dorado, J. R. Rabunal, and A. Pazos, "Automatic epileptic seizure detection in eegs based on line length feature and artificial neural networks," *Journal of neuroscience methods*, vol. 191, no. 1, pp. 101–109, 2010.

[25] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *Journal of neuroscience methods*, vol. 193, no. 1, pp. 156–163, 2010.

[26] K. C. Tan, Q. Yu, C. Heng, and T. H. Lee, "Evolutionary computing for knowledge discovery in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 27, no. 2, pp. 129–154, 2003.

[27] M. Castelli, L. Vanneschi, and S. Silva, "Semantic search-based genetic programming and the effect of intron deletion," *Cybernetics, IEEE Transactions on*, vol. 44, no. 1, pp. 103–113, 2014.

[28] M. Zhang, X. Gao, and W. Lou, "A new crossover operator in genetic programming for object classification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 5, pp. 1332–1343, 2007.

[29] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[30] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.