

# **Novel Genetic Algorithm for Highly Accurate Data Classification using Evolutionary Decision Tree**

***Chuan (Kevin) Huang***  
***Johns Hopkins University, Santa Clara, CA, USA***

[chuang37@jhu.edu](mailto:chuang37@jhu.edu)

## **Abstract**

In order to accurately and efficiently process large and noisy datasets, I designed and implemented a novel algorithm called Direct Encoding Genetic Algorithm (DEGA). It incorporates a novel encoding and decoding algorithm, a unique fitness function, random selection function with proportional crossover mutations, and replacement strategies. For every genotype, a unique strand of genotype is generated with each gene randomly selected from a large genetic pool to prevent repeats. During decoding, the modulus of the gene against the number of remaining features determines the next available feature at every split of the decision tree regardless of neighboring decision information. This ensured a large pool of genetic material that tracks valuable classification information without the fear of rapid convergence.

The results of the algorithm tested against noisy datasets from the University of California – Irvine machine learning database showed near 100% correct classification of the dataset using only 10% of the data for cross validation training. After reducing the amount of training data to 5%, the results still showed 99% correct classification for all datasets. The results showed that the DEGA is able to accurately and efficiently process large and noisy datasets.

**Keywords:** Genetic Algorithm, Direct Encoding Genetic Algorithm, DEGA

## **Biography**



**Kevin Huang** recently graduated from the Johns Hopkins University with a MSE in Computer Science. During his time as a graduate student, Kevin worked on cutting edge problems involving large and often noisy datasets. Through continuous dedication, Kevin was able to create the innovative genetic algorithm using evolutionary decision tree outlined in this research abstract.

His graduate thesis involved creating a brand new algorithm called 3DART for the effective segmentation, identification, and tracking of axoplasmic reticula in the electron microscopy images of the brain. This effort provided significant help in creating seeds and forming connectomes for the Open Connectomes Project.