Abstract

In this research work, we built and ensembled different EGFR microdeletion mutations' based Artificial Neural Networks (ANNs) for improved diagnosis of Non-Small Cell Lung Cancer (NSCLC). We developed two novel algorithms, namely; Genomic Nucleotide Encoding & Normalization (GNEN) algorithm to encode and normalize the EGFR nucleotides and
SimMicrodel algorithm to programatically simulate microdeletion mutations. Sample patients’ data with microdeletion mutations were extracted from online EGFR mutation databases and the two novel algorithms (implemented in MATLAB) were applied to these data to generate appropriate data sets for training and testing of the networks.

The networks after proper training, were combined using minimum error voting ensembling to predict the number of nucleotide deletions in NSCLC patients. Using this ensembling approach, our simulations achieved predictions with minimal error and provides a basis for diagnosing NSCLC patients using genomics based ANN.

**Reference**

- Genetics of Cancer, www.britannica.com
- Eunice L. K., Janusz J., Sarah P. T. Epidermal Growth Factor Receptor Kinase Domain Mutations in Esophageal and Pancreatic Adenocarcinomas, www.aacrjournals.org
- Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen. Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles. National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China.

**Index Terms**

Computer Science Biomedical

**Key words**

ANN EGFR GNEN

NSCLC LM SimMicrodel