Notification No: /547/2023
Date of Award: 17/10/2023

Name of Scholar: Ahmad Abu Turab Naqvi

Name of Supervisor: Prof. Syed Afzal Murtaza Rizvi

Name of Department: Computer Science

Topic of Research: "Analysis of Gene Expression Data for Cancer Classification using Artificial Intelligence Techniques"

Findings

The global incidence of cancer is projected to rise to 25 million cases annually, resulting in around 13 million fatalities by 2030. Cancer classification involves identifying and categorizing cancers based on various attributes such as origin, histological characteristics, genetic modifications, and molecular signatures. This study aims to address the need for a more precise and robust cancer classification approach using artificial intelligence techniques, specifically focusing on gene expression data.

The significance of classification in the development of effective treatment strategies cannot be overstated, but current methods are limited in their capacity to fully comprehend the intricacy and diversity of cancer. Artificial intelligence has proven to be highly effective in various domains of cancer research, leveraging cancer genomics data to enhance disease prediction.

The present study employed downstream analyses of gene expression data derived from diverse cancer types to identify biologically significant genes that exert crucial functions in cellular developmental pathways. Machine learning algorithms, including logistic regression, support vector machine, random forest, decision tree, and gradient boosting, were employed to construct classification models that effectively differentiate between cancerous and noncancerous samples by analyzing gene expression patterns.

The suggested methodology for analyzing gene expression data from eight Chromobox family genes in a Pan-Cancer context yielded statistically significant outcomes. Chromobox genes exhibit notable alterations in expression levels within cholangiocarcinoma, colon adenocarcinoma, lung adenocarcinoma, and lung squamous cell carcinoma. The expression of the CBX2 gene has been observed to be significantly elevated in various types of cancer,

suggesting its potential as an oncogene. Additionally, it was observed that the CBX6 and CBX7 genes exhibited downregulation, suggesting their involvement in tumor suppression. Machine learning algorithms, including logistic regression, support vector machine, random forest, gradient boosting, and decision tree, were utilized to construct classification models that demonstrated a notable ability to accurately differentiate between tumor and normal samples. The analysis of differential expression of the CBX gene in glioblastoma multiforme demonstrated a notable upregulation of these genes in tumor samples in comparison to adjacent normal samples. The analysis conducted using cox regression proportional hazard revealed a noteworthy association between modified gene expression and unfavorable patient overall survival, indicating a significant correlation. The performance evaluation of machine learning algorithms employed for binary classification of tumor and normal samples involved the utilization of several metrics, namely the ROC AUC curve, learning curve, confusion matrix, and mean squared error. All the models exhibit a prediction accuracy of 100%, except for the logistic regression model, which demonstrates a prediction accuracy of 98%, a statistically significant result. The analysis of gene expression in cholangiocarcinoma has demonstrated significant dysregulation of Polycomb group (PcG) genes in tumor samples. In the context of cholangiocarcinoma, it was observed that out of the 33 genes examined through differential expression analysis, a significant alteration in mRNA expression levels was detected in 26 of these genes when comparing cholangiocarcinoma samples to adjacent normal samples. Fourteen genes belonging to the Polycomb Group (PcG) exhibit log foldchange values exceeding two, suggesting significant upregulation. The results of the functional enrichment analysis indicate that the differentially expressed genes are involved in various developmental pathways, such as heterochromatin modulation, stem cell pluripotency, histone modification, and others. The PcG gene is known to exert a significant influence on cellular development and proliferation. The potential consequence of gene dysregulation is the manifestation of aberrant cellular functioning. Consequently, it is imperative to conduct experimental studies to examine their behavior in the context of cholangiocarcinoma.

Prof. & Head
Department of Computer Science
Jamia Millia Islamia
New Delhi-110025