Identifying Subtypes of Cancer Using Genomic Data by Applying Data Mining Techniques

Tejal Upadhyay, Nirma University, Ahmedabad, India
Samir Patel, Pandit Deendayal Petroleum University, Gandhinagar, India

ABSTRACT

This article is about the study of genomics structures and identifying cancer types from it. It divides into six parts. The first part is about the introduction of cancer, types of cancers, how cancer arises, etc. The second part is about the genomic study and how cancer is related to that, which features are used for the study. The third part is about the software which the authors have used to study these genomic structures, which data sets are used, and what is the final output for this study. The fourth part shows the proposed algorithm for the study. The fifth part shows the data preprocessing and clustering. Different preprocessing and clustering algorithms are used. The sixth part shows the results and conclusion with a future scope. The genomics data which is used by this article is taken from the Cancer Genome Atlas data portal which is freely available. Some applied imputation techniques fill up for the missing values and important features are extracted. Different clustering algorithms are applied on genome dataset and results are generated.

KEYWORDS

Cancer, Clustering, Gene Expression Microarray, Genome

1. INTRODUCTION

Cancer is the disorder of body chemicals. In a wide range of disease, a portion of the body’s cells start to separate ceaselessly and spread into encompassing tissues. Ordinarily human cells develop and partition to frame another cell as the body needs them. When cells grow abnormally and cells are added unnecessarily. That cells produced tumor, tumor in human body is either benign or malignant. Malignant tumors are spread into other part of body while benign tumors are sometimes be quite large. When we remove benign cells, they usually do not grow back while malignant tumor may occur recurrently.

Normal cells are dissent from cancer cells that they develop crazy and find yourself intrusive malignancy cells might need the capability to impact the everyday cells, particles and veins that comprehend and feed a tumor a territory called the microenvironment. Malignancy cells will instigate close-by standard cells to form veins that provide tumors with element and supplements, that they need to develop. These veins likewise expel squander things from tumors. Malignant growth is caused
by specific changes to qualities, the essential physical units of heritage. Genes are organized in long stands of tightly packed chemical compound named as chromosomes. The main three types of genetic changes may occur in developing cancer cells:

1. Photo-oncogenes;
2. Tumor suppressor genes;
3. DNA repair genes.

Photograph qualities are including in old cell development and division. When these qualities are modified in positive manners by which amid which they will progress toward becoming malignant growth dispensing qualities. Development silencer are additionally associated with prevailing cell development and division cells with positive changes in development. Silencer qualities may isolate in scholarily degree uncontrolled way. Polymer fix qualities are engaged with settling broken polymer. Cells with changes in these qualities will in general build up extra transformations in these qualities will in general build up extra transformations in different qualities. Together, these changes may make the cells wind up carcinogenic.

Malignant growth is caused by positive changes to qualities, the fundamental physical units of legacy. Qualities are sorted out in long strands of firmly pressed polymer saw as chromosomes. In metastasis, cancer cells break free from wherever they initial shaped (primary cancer), travel through the blood or liquid body substance system, and kind new tumors (metastatic tumors) in alternative elements of the body. The tumor l malignant tumor l malignant neoplasm l tumor l tumor l neoplasm l malignancy l malignance} is that the same kind of cancer because the primary tumor.

Normal cells could become cancer cells. Before cancer cells type in tissues of the body, the cells bear abnormal changes referred to as dysplasia and abnormally. In dysplasia, there’s a rise within the range of cells in an organ or tissue that seem traditional beneath a magnifier. In abnormally, the cells look abnormal beneath a magnifier however don’t seem to be cancer. Dysplasia and abnormally could or might not become cancer.

Our previous work (Upadhyay & Patel, 2018) is extended in this article. Previously we have taken some samples of Blood cancer Leukemia and applied super vised learning classification algorithms. In this article we have taken some genome datasets and applied unsupervised learning clustering algorithms.

2. GENOMIC STUDY RELATED TO CANCER

Any genomic information set involving quite one information kind measured within the same set of tumors as multiple genomic platform (MGP) data (Shen, Ronglai, Olshen, & Ladanyi, 2009). The human ordination sequence is a very important milestone in life science distinctive all the genes associated their restrictive regions provides the essential framework for the genetic blueprint of human race and can facilitate an understanding of the molecular basis of illness.

Technological advances, combined with the event of omic approaches (i.e., genomics, metabolomics and proteomics) the excellent perspective addresses the identification, characterization and quantitation of the full repertoire of genes, proteins and metabolites occurring in living organisms (Casado-Vela, Cebrián, Gómez del Pulgar, & Lacal, 2011).

The Cancer Genome Atlas (TCGA) was launched in 2005 by NIH’s National Human Genome Research Institute (NHGRI) and National Cancer Institute (NCI) to map the key genomic changes in 33 types of cancer. The multi-institution collaboration focused not only on cancer genome sequencing, but also on different types of molecular data collection and analysis, such as investigating gene and protein expression profiles (when they are turned on or off) and associating them with clinical and imaging data. With over $300 million in total funding, the project involved more than 150 researchers at more than two dozen institutions (https://www.cancer.gov).
A Dynamic Agent-Based Model of Corruption
www.igi-global.com/chapter/dynamic-agent-based-model-corruption/21124?camid=4v1a

Evolutionary Algorithms
www.igi-global.com/chapter/evolutionary-algorithms/21119?camid=4v1a