iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5045

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Classification of Cancer using Machine Learning

Kushagra Singh[1], Indrajeet Chatterjee[2], Himanshu Singh[3]
[1, 2, 3]*B.Tech, Dept. of Computer Science, Raj Kumar Goel Institute Of Technology, Ghaziabad, India*

Abstract: ***The trendy genetic code is traditionally represented as an mRNA codon due to the fact, while proteins are made in a cell with the aid of ribosomes, it's far mRNA that directs protein synthesis. The mRNA collection is determined through the collection of genomic DNA. With the rise of computational biology and genomics, most genes are now determined at the DNA level, so a DNA codon is becoming an increasing number of useful. The codon ATG each codes for methionine and serves as an initiation site: the first ATG in an mRNA's coding location is wherein translation into protein starts off evolved. The other start codons indexed by way of Gene Bank are uncommon in eukaryotes and normally codes for Met/F-Met. Our method for extracting time established facts about a person mutation history is to focus at the tandem repeat areas of this individual healthful genome. The repetition of a subsequence is (TCAT -> TCATCATG). From this sequence of DNA the image file will be processed and classified into several categories dependent on their past environment where the particular person livelihood would be. We have used Jupyter notebook and KAGGLE data set for classification using linear regression algorithm. This project will be useful for predicting cancer in distinguish part of the world. This will be very optimal and consistent for diagnosis. It can be improved with passage of its utility while monitoring several issues related with cancer such as RBC minimization and WBC maximization.***

I. INTRODUCTION

Over the past few decades, there has been much research on diagnosing cancer at very early stages. Cancer is a group of disease which involves abnormal cell growth in our body. There are many types of cancer like Breast cancer, Prostate cancer, Skin cancer, Colon cancer, etc. Diagnosis of cancer at an early stage is a must to have a better treatment. In the modern era with an increase in technology and computational power of machine a lot of data regarding various types of cancer is been collected and is used for various medical research purposes. Even with an increase in technology one of the most challenging task is to predict cancer with good accuracy. To achieve good accuracy many researchers are using various Machine learning algorithms. The Use of various ML methods is helping various researchers and scientists are able to predict cancer and can identify various patterns in DNA of people and the relationship between the pattern of DNA and type of cancer. Still there is a long way to go in order to predict cancer with 100% accuracy but there is no doubt using ML methods has increased the chance of predicting cancer at an early stage and allowing proper time for the patient to carry his/her treatment.

II. OBJECTIVE

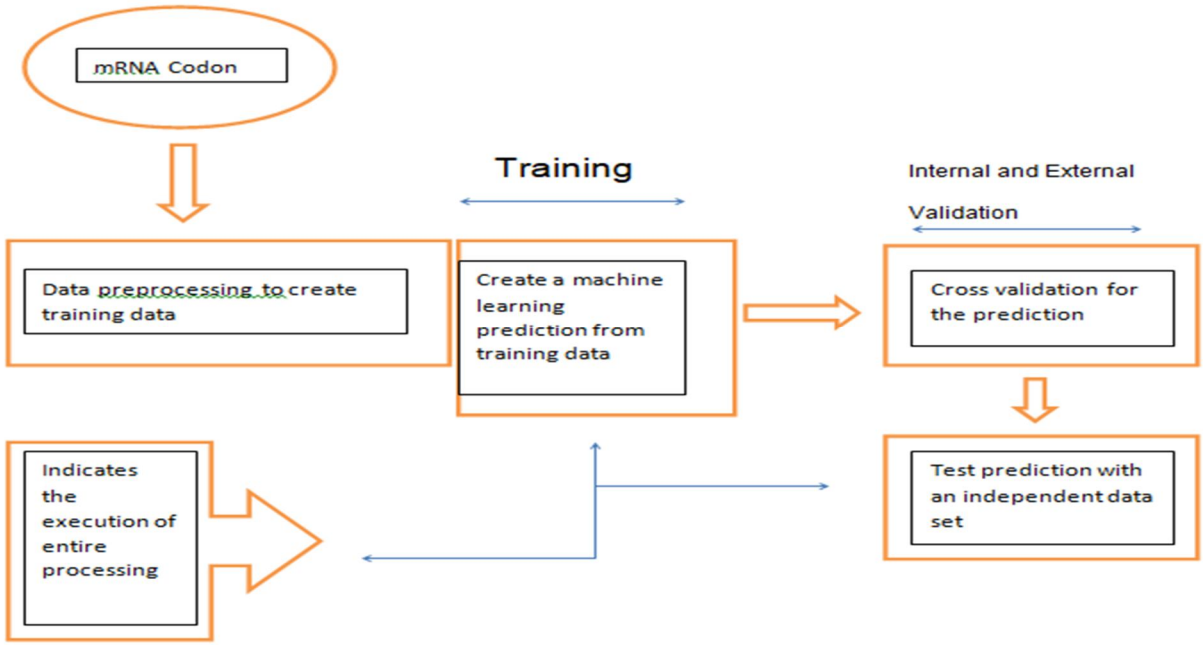
The main objective is to study the various pattern of mRNA in order to predict that particular person will have cancer or not. The pattern could be of type(TCAT->TCATCATG). In our project we are using 3842 blood derived samples of 11 cancers for the analysis. As we have seen the mutation of blood derived from normal DNA of patients which shows strong signals from rest of tested cancers with accuracies ranging in between 76% to as high as 92%. Cancers which emit different mutation are easier to distinguish i.e in more accurate classifiers. There are different type of cancer on the basis of pattern of DNA which led us to define four different cancer classes i.e Class 1 = GBM, Class 2 = SKCM, Class 3 = PAAD and Class 4 = HNSC, BLCA, LGG, LUAD, LUSC, PRAD, STAD, THCA.

III. LITERATURE SURVEY AND DRAWBACKS

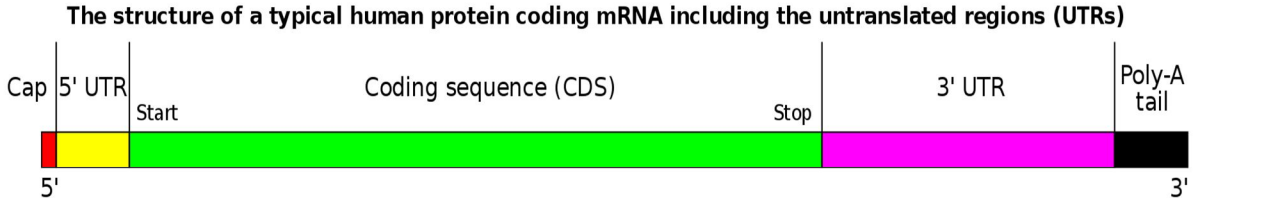
Title	Description	Drawbacks
The classification of cancer using DNA micro array.	They compare the raw DNA with aged DNA and tried to classify the heredity feature of cancer which could be improved in future in a particular person.	It takes one person entire life in order to complete the thesis. It is very crucial that each and every moment gets captured for the improvement of the process.
The polymorphism of cancer through neural networks and its classification.	A GP based method gene rating DNA models for cancer classification named as 7129. It is signed in precession values 60 patients of whom 39 survived.	The over fitting is a dominant cancer with ML approaches to DNA chip data. GP is a stochastic process which runs at least 10 times and is a topic of discussion.
The prediction of type of cancer using DNA signature.	It uses 6640 Tumor sample and 28 cancer type cell using ML: Linear Support Vector machine regularized logistic and random observation.	It deals with only 28 cancer type cells where there are more than 100 types of cancer cell are there. The accuracy rate is only 49.4 i.e. 0.4%

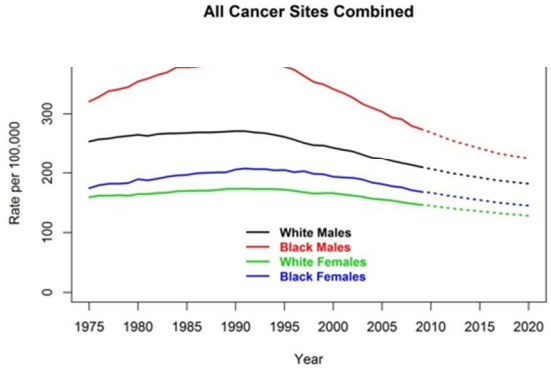
©IJRASET: All Rights are Reserved

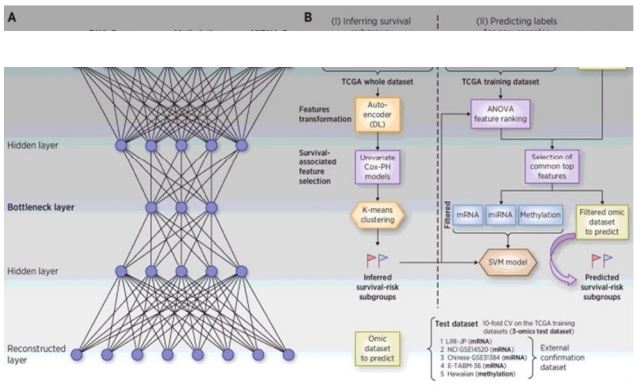
A. Model Making



B. Module Figure







C. Proposed System

This system is highly improvised from the previous systems with respective technology and features. The main aim is to reduce the processing time and make the model efficient for handling all sorts of parameters related to cancer.

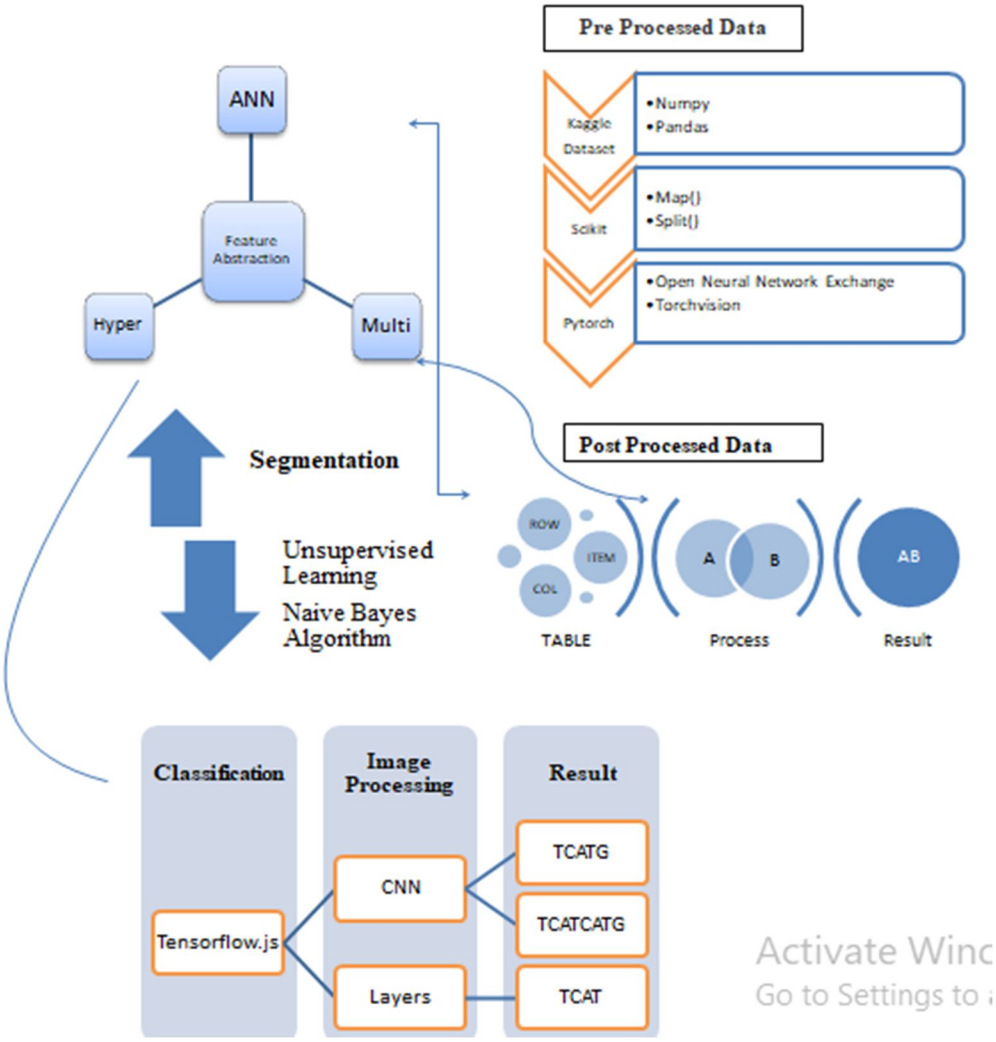
Earlier	Now
HNSC 65% Accuracy	HNSC 71% Accuracy
SKCM 85% Accuracy	SKCM 93% Accuracy
PAAD 52% Accuracy	PAAD 71% Accuracy

©IJRASET: All Rights are Reserved

1) HNSC: HNSC-HYDROSAP The hydrogen molecule that present in mRNA which signifies the stability.

Naïve Bayes algorithm form the layers of CNN which basically helps in better segmentation and visualization process. Further it helps in training the entire model.

D. Architecture



IV. CONCLUSION

To achieve good performance for predicting cancer in the most diversified structure which includes place, livelihood, diet, surrounding and particular mRNA, we used SKCM and PAAD data set which we further processed using Kerras framework and therefore we are able to optimized the data set, for further processing where we used Un-Supervised learning in segmentation and Naïve Bayes algorithm for preparing the exact model configuration. For analyzing and error reducing we used three frameworks inside Kerras framework. We used Numpy library and Pandas library for reducing the unwanted data and making it for optimal and fit for making the effective model. ScikitLearn library for establishing the logarithmic relation in the particular dataset which comes with matplot.lib function for making graph and split function for splitting the data into specific order. And Pytorch framework is being used for detecting invisible patches in the developed model. The final result comes from Tensorflow.js classifier which is the most powerful classifier framework available and is being used in the industry. Tensorflow.js helps in forming the most efficient tensor which is nothing but the relevant rows and columns of our CSV file. It compares with outsources available in the database and therefore improve the process of classification of cancer using mRNA genetic codons. At last we assure that our model can improvised the cancer classification process in the most relevant way since we have used the diversified structure and can provide exact probability of having a cancer or not which tends to be accurate.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Qingzhong Liu1, Andrew H Sung2*, Zhongxue Chen3, Jianzhong Liu4, Lei Chen1, Mengyu Qiao5, Zhaohui Wang6, Xudong Huang7, Youping Deng6,8*From BIOCOMP 2010. The 2010 International Conference on Bioinformatics and Computational Biology Las Vegas, NV, USA. 12-15 July 2010: Gene selection and classification for cancer microarray data based on machine learning and similarity measures.

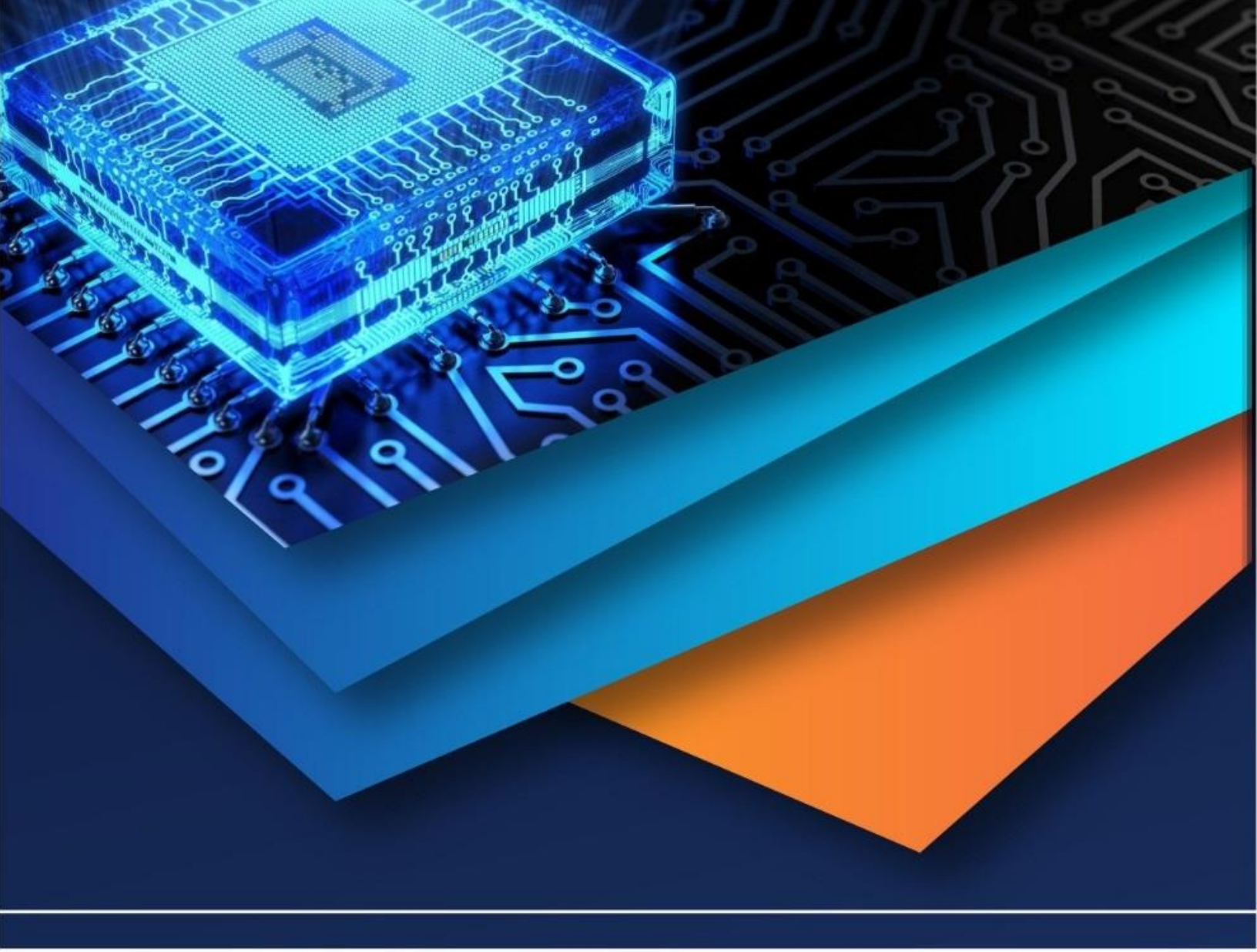
[2] Yang, S. and Naiman, D., 2014. Multiclass cancer classification based on gene expression comparison. *Statistical Applications in Genetics and Molecular Biology*, 0(0).

[3] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzisc, Dimitrios I. Fotiadis.: Machine learning applications in cancer prognosis and prediction

[4] Alireza Osareh, Bita Shadgar:Machine Learning Technique to diagnose Breast Cancer.

[5] Andrew Dave, YashBhatiya: Cancer prediction using segmentation of blood cell.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)