WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI Technique for Staging Cancer

Sreekarthik C.P[1], Prof Rajitha P.R[2], Dr.T.mahalekshmi[3]

[1]Final year Student, Sree Narayana Instiute of Technology, Kollam, Kerala,India
[2]Assistant Professor, Sree Narayana Instiute of Technology, Kollam, Kerala,India
[3]Principal, Sree Narayana Instiute of Technology, Kollam, Kerala,India

ABSTRACT:

Cancer is one of the world's most dreaded and deadly diseases, accounting for more than 9 million deaths worldwide. Early detection of cancer improves the odds of a successful recovery. RNA sequence analysis is one staging technique. The study of human genetics has been made easier thanks to recent developments in the efficiency and accuracy of artificial intelligence approaches and optimization algorithms. To diagnose different forms of cancer based on tumour RNA sequence (RNA-Seq) gene expression data, this research offers a unique optimised deep learning strategy based on binary particle swarm optimization with decision tree (BPSO-DT) and convolutional neural network (CNN). Kidney renal clear cell carcinoma (KIRC), breast invasive carcinoma (BRCA), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), and uterine corpus endometrial carcinoma will all be studied in this study (UCEC). The proposed strategy is divided into three stages.The first phase is preprocessing, which involves utilising BPSO-DT to optimise the high-dimensional RNA-seq and then converting the improved RNA-seq to 2D pictures. The augmentation phase raises the original dataset of 2086 samples to 5 times its original size. The augmentation approaches were chosen with the goal of having the least amount of impact on the image's attributes. This phase aids in overcoming the problem of overfitting and trains the model to improve accuracy. Deep CNN architecture is the third step. This phase introduces an architecture consisting of two major convolutional layers for feature extraction and two fully connected layers to categorise the five different types of cancer based on the photos available on the dataset.

INTRODUCTION:

Cancer is a broad term that refers to a set of disorders characterised by aberrant cell proliferation, metastatic spread, and invasiveness. To diagnose different forms of cancer based on tumour RNA sequence (RNA-Seq) gene expression data, this research offers a unique optimised deep learning strategy based on binary particle swarm optimization with decision tree (BPSO-DT) and convolutional neural network (CNN). The high-dimensional RNA-seq data was optimised with BPSO-DT to minimise its dimensions by picking only the best features and deleting the irrelevant features, resulting in a high level of classification accuracy. The refined RNA-seq results were then incorporated into 2-D pictures to feed into the CNN architecture.Different data augmentation techniques have been applied to the 2D photos to avoid overfitting. The proposed method was trained and tested on a publicly available RNA-seq dataset that included five cancer types: kidney renal clear cell carcinoma (KIRC), uterine corpus endometrial carcinoma (UCEC), breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), and lung squamous cell carcinoma (LUSC) (LUSC).

TECHNOLOGIES:

METHODOLOGY

1) PRE-PROCESSING PHASE (BPSO-DT AND 2D IMAGE CREATION)

   BPSO is utilised to implement feature selection in this phase, with the decision tree (DT) serving as BPSO's fitness function for a classification problem. BPSO is utilised in this study to decrease the amount of RNA-seq features to a minimum, choose only the most significant characterstics and improve classification accuracy.

   The top features of RNA-seq were chosen from a total of 971 characteristics in the BPSO-DT processing. The refined tumour gene expression dataset was then subjected to a series of processes to convert it from data to picture format. 1) Load the tumour gene expression into memory, 2) Change the data numerical domain range from [0, 24248] to image domain range [0, 255], and 3) Construct image by converting the optimal data record of 615 cells into a 25 x 25 pixels image. This phase will yield 2086 images that will be sorted into five tumour groups. The following is the equation to convert to an image:

   Where 24248 is the maximum cell value in the tumour gene expression data and 255 is the maximum value of the image domain,

PixelValue = Round (Cell Value 255 / 24248)

2) DATA AUGMENTATION PHASE

After preprocessing, the original dataset contains 2086 pictures for five classes of tumour gene expression. The model is very likely to overfit because to the large disparity between learnable parameters and the amount of photos in the training set. With large datasets, deep learning models perform better. Data augmentation or jittering is a common approach to increase the size of datasets. When training on very little data, data augmentation can expand the size of the dataset by up to 10 or 20 times or more. This helps minimise overfitting. The most popular way to avoid overfitting is to utilise label-preserving modifications to increase the number of images used for training. Furthermore, data augmentation strategies are employed to the training set in order to make the resulting model more resistant to reflection, zooming, and tiny noise in pixel values. Each image in the training data is used to perform augmentation.

- Reflection around X axis,
- Reflection around Y axis,
- Reflection around X-Y axis,
- Zooming.

3) DEEP LEARNING TRAINING PHASE

The CNN model is trained to classify the five different forms of cancer in this phase. Two convolutional layers for feature extraction with varying convolution window 3*3 pixels, followed by two fully connected layers for classification, make up the architecture. The input layer is the initial layer, having a 25*25 pixel input size. The convolution layer, with a window size of 3*3 pixels and 16 distinct filters, is the second layer. The nonlinear activation function is a ReLU in the third layer, which is followed by an intermediate pooling with subsampling in layer four. The sixth and seventh layers use a convolution layer with a window size of 3*3 pixels, 32 distinct filters, and the ReLU activation function. To avoid the problem of overfitting, layer number eight contains a dropout layer.Then there's layer nine, which is a completely connected layer with 64 neurons and ReLU activation. In layer thirteen, the final fully linked layer uses a softmax layer to get class memberships and has 5 neurons to classify 5 classes for tumour gene expression.

Existing System

Many traditional methods for diagnosing cancer type are employed, but these procedures may take longer to find the tumour type. Imaging, laboratory tests (including tumour marker tests), tumour biopsy, endoscopic examination, surgery, and genetic testing are all possible cancer diagnostic techniques. As a result, these treatments take a long time to complete.

Disadvantages of existing system

Some of the disadvantages of the existing system are as follows:

- ❖ There is no mechanism to predict the cancer type accurately.
- ❖ Existing systems takes more time to find the type.
- ❖ Sometimes this system may cause failures and giving false results.

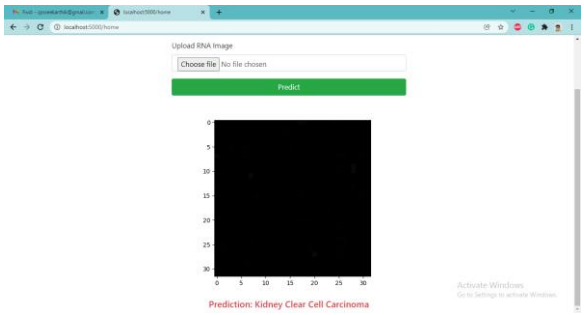
Proposed System

- An optimised deep learning strategy based on BPSO-DT and CNN is used in the proposed system to categorise normal and malignant conditions based on high-dimensional RNA-Seq gene expression data.
- Kidney renal clear cell carcinoma (KIRC), breast invasive carcinoma (BRCA), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), and uterine corpus endometrial carcinoma will all be studied in this study (UCEC).
- This creates a user-friendly interface, avoiding potentially perplexing instances.

Advantages

1. Works very fast.
2. Accurate prediction.
3. User friendly.

RESULTS



CONCLUSION AND FUTURE WORKS

Cancer is a term used to describe a collection of diseases characterised by abnormal cell proliferation and the ability to invade or spread to numerous regions of the human body. More than 9 million people die each year as a result of this disease. Because of increases in efficiency and accuracy, RNA-Seq has substantially boosted the analysis of human genetics, which aids in understanding the nature of cancer disorders. Breast Invasive Carcinoma (BRCA), Kidney Renal Clear Well Carcinoma (KIRC), Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), and Uterine Corpus Endometrial Carcinoma were all classified differently in this paper (UCEC).The proposed strategy was divided into three stages. The first phase is pre-processing, which includes selecting the best features of RNA-Seq and converting them to 2D pictures using the binary particle swarm optimization with design trees (BPSO-DT) technique. The data augmentation process boosted the original dataset volume to 5 times its original size. The deep convolutional neural network architecture is the third step. In this phase, an architecture consisting of two convolutional layers for feature extraction and two fully connected layers to classify the five different cancer kinds was presented. According to the reported data and performance measures used in this study, the proposed method achieved an overall testing accuracy of 96.90 percent. The comparative results were presented, and the accuracy reached in this study beats that of other relevant studies for testing accuracy for five tumour classes.Furthermore, the proposed approach is less difficult and requires less training time. Applying novel topologies of deep neural networks, such as Generative Adversarial Neural networks, is one of the promising future efforts.

Reference

Bi, W.L., Hosny, A., Schabath, M.B., Giger, M.L., Birkbak, N.J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I.F. and Mak, R.H., 2019. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: a cancer journal for clinicians*, *69*(2), pp.127-157.

Niu, P.H., Zhao, L.L., Wu, H.L., Zhao, D.B. and Chen, Y.T., 2020. Artificial intelligence in gastric cancer: Application and future perspectives. *World Journal of Gastroenterology*, *26*(36), p.5408.