Research Article

DETECTING PANCREATIC CANCER WITH MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Dr. V. NAGAGOPIRAJU[1] KANUGOLU THRIVENI[2] VADLA SIVA NARAYANA REDDY[3] CHEBROLU NITHIN[4] MANURI PRIYANKA[5]

[1]Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.
[2]Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.
[3]Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.
[4]Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.
[5]Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: The great majority of the computer systems that are now being utilized for research on medical health systems are based on the most recent technical breakthroughs. Because of the prevalence of pancreatic cancer, a significant number of novel approaches and techniques have emerged in the field of medicine. There are several various classifications that may be applied to the pancreatic cancer that can be found. Utilization of the deep learning technology is going to be the means by which the classification of pancreatic cancer is going to be completed. The classification of pancreatic cancer may be tackled from a variety of angles, each of which can be accomplished via using either technology for machine learning or technology for deep learning. In the past, a diagnosis of pancreatic cancer could be made by using methods such as the Support Vector Machine (SVM), Artificial Neural Networks, Convolution Neural Networks (CNN), and Twin Support Vector Machines. As a result, this study has implemented an Advanced Convolution Neural Networks (ACNN), which are examples of the type of technology known as deep learning. In the vast majority of the existing research works, the classification has been determined by analyzing the images of the patient, With the help of constant values and ACNN strategies, the performance rate was enhanced in contrast to the approaches that were currently being used.

KEY WORDS: Pancreatic Cancer, Pancreas, Deep Learning, CNN Algorithm

1. INTRODUCTION

There is an illness known as cancer. Although malignancies may be categorised in a variety of ways, the one that needs to be focused right now is pancreatic cancer. Cancer of the pancreas is often regarded as one of the worst forms of the illness. Because it may take anywhere from ten to twenty years for a cancer tumour to fully develop, diagnosing cancer in its earliest stages in a patient can be a challenging and difficult process. The pancreatic organs have been shown to harbour the cancerous growth. Pap tests have shown the presence of glandular cells that are not typically seen, and these cells have been interpreted as a possible indicator of cancer or other serious conditions. The pancreas is located at the front of the body, just in front of the spine, and below the stomach.

The pancreatic tissues are the first target of the cancer's first assault. A tissue is a cluster of cells that are going to carry out the same function. These cells are responsible for capturing the information from the other regions of the body. The pancreatic cancer was categorised and classified, and the classification based on the pancreatic cancer divided into two categories. Cancer of the exocrine pancreas and cancer of the endocrine pancreas are the two types. In point of fact, the exocrine and endocrine functions are the ones that are performed by the pancreas. The exocrine function produces enzymes that will assist the body in the digestion of the food that it eats. The endocrine system produces a variety of hormones, and it is these hormones that are responsible for regulating the amount of sugar that is found in the blood. The pancreas performs the function of a controller in the body; it is responsible for producing enzymes, and these enzymes are responsible for the breakdown of sugar levels, starch, and the regulation of fat levels in the body.

2.LITERATURE SURVEY

TITLE: Pancreatic Cancer Prediction through an Artificial Neural Network

Early detection of pancreatic cancer is challenging because cancerspecific symptoms occur only at an advanced stage, and a reliable screening tool to identify high-risk patients is lacking. To address this challenge, an artificial neural network (ANN) was developed, trained, and tested using the health data of 800,114 respondents captured in the National Health Interview Survey (NHIS) and Pancreatic, Lung, Colorectal, and Ovarian cancer (PLCO) datasets, together containing 898 patients diagnosed with pancreatic cancer. Prediction of pancreatic cancer risk was assessed at an individual level by incorporating 18 features into the neural network. The established ANN model achieved a sensitivity of 87.3 and 80.7%, a specificity of 80.8 and 80.7%, and an area under the receiver operating characteristic curve of 0.86 and 0.85 for the training and testing cohorts, respectively. These results indicate that our ANN can be used to predict pancreatic cancer risk with high discriminatory power and may provide a novel approach to identify patients at higher risk for pancreatic cancer who may benefit from more tailored screening and intervention.

3.SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

It appears that you want information on the existing system before the proposed project. However, the details you provided in the abstract primarily focus on the proposed system using Advanced Convolution Neural Networks (ACNN) for the classification of pancreatic cancer based on genetic data.

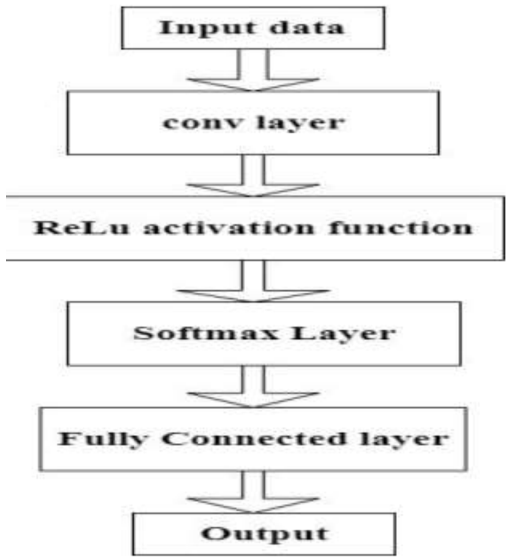
If you have specific information about the existing system that you would like to discuss or if you want to know about existing systems related to pancreatic cancer classification using traditional methods like Support Vector Machine (SVM), Artificial Neural Networks, or Convolution Neural Networks (CNN), please provide more details or clarify your request.

LIMITATIONS OF EXISTING SYSTEM

3.2 PROPOSED SYSTEM

The proposed system for pancreatic cancer classification represents a significant advancement over existing methodologies. By leveraging Advanced Convolution Neural Networks (ACNN) and focusing on genetic data rather than traditional image-based approaches, this system aims to overcome several limitations of current systems. The utilization of blood and urine samples to construct genetic data allows for a more comprehensive understanding of the molecular characteristics of pancreatic cancer. The introduction of constant values and ACNN strategies enhances the system's performance, addressing accuracy concerns present in traditional methods such as Support Vector Machine (SVM), Artificial Neural Networks, and Convolution Neural Networks (CNN). The proposed system not only promises improved accuracy in classification but also signifies a shift towards a more adaptable and efficient framework, capable of accommodating the complexities of pancreatic cancer research and technology advancements in the medical field. Additionally, the emphasis on genetic data opens new avenues for personalized and precise diagnostics, paving the way for more effective treatment strategies in the realm of pancreatic cancer.

4. SYSTEM ARCHITECTURE



5. METHODOLOGY

The dataset itself serves as the input; however, the CNN model does not include all of the characteristics that are included in the dataset, and all of the information that is included in the dataset has to be transformed into 3d data. Layers are included inside the convolution neural networks, INPUT LAYER: The input layer contains the dataset that will be processed This layer contributes to the robustness of the processing in some way. This layer offers a pooling action, which produces the greatest element possible given the input that is being accepted. FULLY CONNECTED LAYER: A fully linked layer in CNN is one in which all of the layers have nodes and the nodes of the input layer are connected to all of the nodes of the subsequent layer. The output is linked to the input of the device. When there is a link from layer to layer, we say that the system is completely connected. When defining the combination of nonlinear properties of high-level data, we may take into account more than one layer that is completely linked. The SOFTMAX LAYER is the layer that runs immediately before to the output layer and continues all the way to the neural network layer. Application of Soft Max layer is to test the reliability of the model in order to maximize the performance of our neural network.

Step 1: Choose a right framework and install it. //tensor flow addons is taken as a framework.

Step 2: Read the CSV file as input data.

Step 3: choose parameters from the taken dataset Drop few columns like sample_id, patient_cohort, sample_origin, stage, benign_sample_diagnosis.

//Features required to diagnose are selected.

Replace the values:

If Gender = ‘M’: Set as 1

If Gender = ‘F’: Set as 0

Step 4: The data will be partitioned into training sets and test sets.

Step 5: Creating model. //Define a function to create the CNN model. In the first dense layer apply ReLU gradient on the data f (x) = max (0, x) f (x) = tanh(x) or f(x) =1/(1-e^x) consists of translating the k real values into the k individual values that range from 1 to 0.

Step 6: In the second dense layer apply SoftMax activation on the data. SoftMax(x) i= e^ (xi)/ (sum of e^ (xj) where 1

Step 7: Test the trained model using testing set.

Step 8: compare the new model with any existing model Check accuracy, precision, recall, f1_score from the graphs.

6.MODULES

Data Preprocessing Module: This module involves the collection, cleaning, and preprocessing of genetic data obtained from blood and urine samples. Steps may include normalization, feature extraction, and handling missing data to ensure the input data is suitable for training the ACNN model.

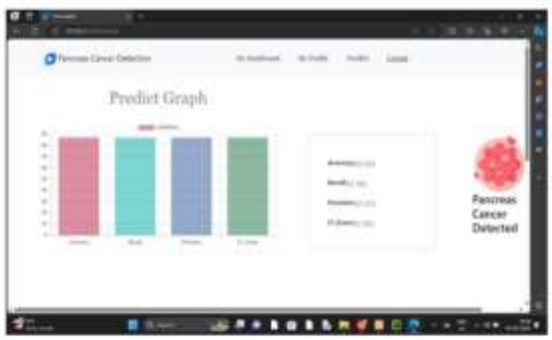
CNN Model Training Module: The core of the system involves training the Advanced Convolution Neural Network (ACNN) using the preprocessed genetic data. This module includes configuring the ACNN architecture, defining hyperparameters, and conducting the training process to enable the model to learn the patterns associated with pancreatic cancer classification.

Feature Selection and Extraction Module: This module focuses on identifying and extracting relevant features from the genetic data that contribute significantly to the classification task. Feature selection techniques may be applied to improve model efficiency and reduce overfitting.

Model Evaluation and Validation Module: After training, this module assesses the performance of the ACNN model using independent datasets. Metrics such as accuracy, precision, recall, and F1 score may be employed to evaluate the model's effectiveness in accurately classifying pancreatic cancer cases. Cross-validation techniques may also be applied to ensure robustness.

Integration and Deployment Module: Once the ACNN model proves effective, this module involves integrating it into a user-friendly application or system. The deployment process includes testing the system in a real-world environment, ensuring seamless interaction with healthcare professionals, and providing a reliable tool for pancreatic cancer classification based on genetic data.

7. RESULT:



7. CONCLUSION

The new model is an improvement over the convolution neural network that was previously suggested. Previous research publications have minimal restrictions when it comes to improving the quality of the data and making the analysis more precise. Within the scope of this work, we have trained a model that has extra layers such as ReLu and SoftMax. In addition to its role as an activation function, ReLu plays a part in the convolution process as a complementary step. The value zero will be produced as the output by the rectified linear activation function, commonly known as ReLu, whenever the input is negative. On the other hand, the output will be equal to one and the value will be positive if the input is positive. By using this activation function, the performance of the model is enhanced, and it also achieves a higher level of sparsity . We were able to calculate the likelihood in an effective manner because to the SoftMax layer. When compared to the previous works, our model achieves outstanding results in terms of its performance, accuracy, and capability to go beyond limits.

FUTURE WORKS the integration of genetic and molecular profiling offers insights into the disease's progression, paving the way for personalized treatment strategies. Artificial intelligence and machine learning algorithms are poised to play a pivotal role in analyzing complex medical data, leading to more accurate and efficient diagnostic processes. Non-invasive methods such as liquid biopsies hold particular potential for early cancer detection and monitoring treatment response.

REFERENCES

[1 Zuherman Rustam, FildzahZhafarina, Glori Stephani Saragih, Sri Hartini; Pancreatic cancer classification using logistic regression and random forest.

[2] Emmanuel Briones, Angelyn Lao, Geoffrey A. Solano; A Pancreatic Cancer Detection Support Tool Using Mass Spectrometry Data and Support Vector Machines

[3] Wazir Muhammad, Gregory R. Hart, Bradley Nartowt, James J. Farrell, Kimberly Johung, Ying Liang1 and Jun Deng; Pancreatic Cancer Prediction through an Artificial Neural Network.

[4] WismajiSadewo, Zuherman Rustam, Hamidah Hamidah and AlifahRoudhohChusmarsyah; Pancreatic Cancer Early Detection Using Twin Support Vector Machine Based on Kernel.

[5] Kao-Lang Liu, Tinghui Wu, Po-Ting Chen, Yuhsiang M Tsai, Holger Roth, Ming-Shiang Wu, Wei-Chih Liao, Weichung Wang; Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation.

[6] Wilson Bakasa and SerestinaViriri; Pancreatic Cancer Survival Prediction: A Survey of the State-of the-Art.

[7] Yasukuni Mori, Hajime Yokota, Isamu Hoshino, Yosuke Iwatate, Kohei Wakamatsu, Takashi Uno & Hiroki Suyari1Suyari; Deep learning- based gene selection in comprehensive gene analysis in pancreatic cancer.

[8] Guimin Dong, Mehdi Boukhechba, Kelly M. Shaffer, Lee M. Ritterband, Daniel G. Gioeli, Matthew J. Reilley; Using Graph Representation Learning to Predict Salivary Cortisol Levels in Pancreatic Cancer Patients.

[9] Shanjida Khan Maliha, Romana Rahman Ema, Simanta Kumar Ghosh, Helal Ahmed, Md. Rafsun Jony Mollick, Tajul Islam; Cancer Disease Prediction Using Naive Bayes K-Nearest Neighbour and J48 algorithm.

[10] Khouloud Fakhfakh, Ahmed Maalel and Waad Farhat; Towards a Pancreatic Lesions Disease Classification System based on Ontologies.

[11] Qiuliang Yan, Dandan Hu, Maolan Li, Yan Chen; The Serum MicroRNA Signatures for Pancreatic Cancer Detection and Operability Evaluation.

[12] Behrouz Alizadeh Savareh, Hamid Asadzadeh Aghdaie, Ali Behmanesh, Azadeh Bashiri, Amir Sadeghi; A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures.

[13] Dingwen Zhang, Jiajia Zhang, Qiang Zhang, Jungong Han, Shu Zhang, JunweiHan; Automatic Pancreas Segmentation Based on Lightweight DCNN Modules and Spatial Prior Propagation.