# Detecting Pancreatic Cancer with Machine Learning and Deep Learning Technique

**[1]Pooja D Y, [2]Siddesh K T, [3]Kotru Swamay S M**

*1, Student, Department Of MCA, BIET, Davangere*

*2, Assistant Professor, Department Of MCA, BIET, Davangere*

*3, Assistant Professor, Department Of MCA, BIET, Davangere*

**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), 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**.**

**Keywords: Pancreatic Cancer, Pancreas, Deep Learning, CNN Algorithm.**

## 1. INTRODUCTION

There is an illness known as cancer. Although malignancies may be categorized 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 totwenty 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 harbor 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 ofthe 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 categorized 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 levelsin the body.

## 2. LITERATURE SURVEY

**TITLE:** Pancreatic Cancer Prediction through an Artificial Neural Network.

Early detection of pancreatic cancer is challenging because cancer specific 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 pancreaticcancer 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 trainingand testing cohorts, respectively.

These results indicate that our ANN can be used to predict pancreatic cancer risk with high discriminatory to identify patients at higher risk for pancreatic cancer who may benefit from more tailored screening and intervention.

## 3. METHODOLOGY

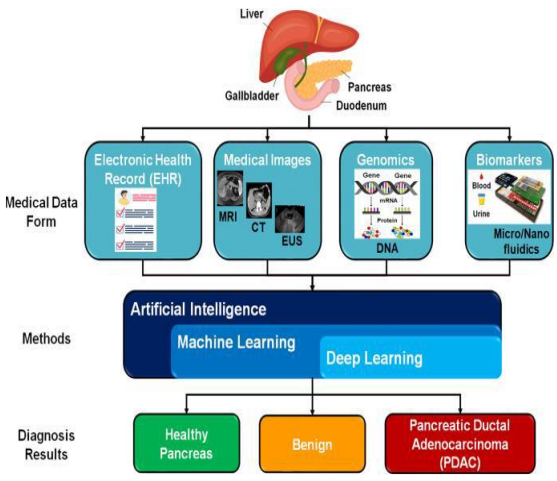


**Figure:3.1 Methodology of pancreatic cancer system**.

The methodology of the Pancreatic Cancer Classification project involves several key steps aimed at developing and validating a robust classification system using Random Forest and Convolutional Neural Network (CNN) models. The methodology can be outlined as follows:

### 3.1 Dataset used:

A dataset containing a diverse dataset of medical imaging data containing pancreatic images, such as CT scans or MRI images, along with corresponding labels indicating cancerous or non- cancerous tissue. In this used to plane having maximum distance of data points.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | sample_ic | patient_c | sample_o | age | sex | diagnosis | stage | benign_sa | plasma_C | creatinine | LVVE1 | REG1B | TFF1 | REG1A |
| 2 | S1 | Cohort1 | BPTB | 33 | F | 1 | | | 11.7 | 1.83222 | 0.893219 | 52.94884 | 654.2822 | 1262 |
| 3 | S10 | Cohort1 | BPTB | 81 | F | 1 | | | | 0.97266 | 2.037585 | 94.46703 | 209.4883 | 228.407 |
| 4 | S100 | Cohort2 | BPTB | 51 | M | 1 | | | 7 | 0.78039 | 0.145589 | 102.366 | 461.141 | |
| 5 | S101 | Cohort2 | BPTB | 61 | M | 1 | | | 8 | 0.70122 | 0.002805 | 60.579 | 142.95 | |
| 6 | S102 | Cohort2 | BPTB | 62 | M | 1 | | | 9 | 0.21489 | 0.00086 | 65.54 | 41.088 | |
| 7 | S103 | Cohort2 | BPTB | 53 | M | 1 | | | | 0.84825 | 0.003393 | 62.126 | 59.793 | |
| 8 | S104 | Cohort2 | BPTB | 70 | M | 1 | | | | 0.62205 | 0.174381 | 152.277 | 117.516 | |
| 9 | S105 | Cohort2 | BPTB | 58 | F | 1 | | | 11 | 0.89349 | 0.003574 | 3.73 | 40.294 | |
| 10 | S106 | Cohort2 | BPTB | 59 | F | 1 | | | | 0.48633 | 0.001945 | 7.021 | 26.782 | |
| 11 | S107 | Cohort2 | BPTB | 56 | F | 1 | | | 24 | 0.61074 | 0.278779 | 83.928 | 19.185 | |
| 12 | S108 | Cohort2 | BPTB | 77 | F | 1 | | | | 0.29406 | 0.001176 | 6.218 | 28.297 | |
| 13 | S109 | Cohort2 | BPTB | 71 | M | 1 | | | 23 | 1.05183 | 0.860337 | 243.082 | 608.284 | |
| 14 | S11 | Cohort1 | BPTB | 49 | F | 1 | | | | 0.85956 | 1.416314 | 151.8308 | 74.1899 | 505.571 |
| 15 | S110 | Cohort2 | BPTB | 53 | M | 1 | | | 7 | 1.91139 | 1.516773 | 150.89 | 590.686 | |
| 16 | S111 | Cohort2 | BPTB | 56 | F | 1 | | | 12 | 0.91611 | 0.599645 | 93.811 | 93.576 | |
| 17 | S112 | Cohort2 | BPTB | 60 | F | 1 | | | 28 | 0.50895 | 0.002036 | 24.366 | 19.698 | |
| 18 | S113 | Cohort2 | BPTB | 69 | F | 1 | | | 9 | 0.41847 | 0.001674 | 17.102 | 0.032641 | |
| 19 | S114 | Cohort2 | BPTB | 60 | F | 1 | | | 47 | 0.80301 | 0.003212 | 3.588 | 30.071 | |
| 20 | S115 | Cohort2 | BPTB | 55 | M | 1 | | | 17 | 1.28934 | 2.285351 | 67.468 | 269.805 | |
| 21 | S116 | Cohort1 | BPTB | 28 | F | 1 | | | 8.7 | 0.50895 | 0.58301 | 13.61906 | 267.1935 | 381 |
| 22 | S117 | Cohort1 | BPTB | 54 | F | 1 | | | | 1.2441 | 0.004976 | 5.50735 | 193.1457 | 113 |
| 23 | S118 | Cohort1 | BPTB | 50 | F | 1 | | | 8.7 | 0.95004 | 0.0038 | 56.39913 | 192.2589 | 137 |
| 24 | S119 | Cohort1 | BPTB | 40 | M | 1 | | | | 0.76908 | 0.653984 | 14.60758 | 341.2675 | |
| 25 | S12 | Cohort1 | BPTB | 74 | F | 1 | | | | 0.31668 | 0.58301 | 25.52035 | 146.5886 | 111.531 |
| 26 | S120 | Cohort1 | BPTB | 63 | M | 1 | | | | 0.75777 | 2.44018 | 21.22948 | 109.4212 | 903 |
| 27 | S121 | Cohort1 | BPTB | 50 | M | 1 | | | | 0.78039 | 1.044411 | 7.355656 | 250.3443 | 149 |

### 3.2 Data Preprocessing:

Preprocess the imaging data, including standardization, normalization and potentially augmentation techniques to enhance data quality and facilitate model training. Clean the collected data by removing noise, irrelevant information, and formatting inconsistencies. Normalize and standardize the data to make it suitable for analysis.

### 3.3Feature Extraction:

Extract relevant features from the preprocessed imaging data that capture distinctive patterns and characteristics indicative of pancreatic cancer. For Random Forest model, feature selection techniques may be employed to identify the most discriminative features. For CNN model, employ convolutional layers to automatically learn hierarchical features directly from the raw pixel data

### 3.4 Model Development and Training:

Implement Random Forest and CNN models using appropriate libraries (e.g., scikit-learn for Random Forest, TensorFlow or PyTorch for CNN). Split the dataset into training, validation, and test sets. Train the models on the training set, optimizing hyperparameters and model architectures through techniques such as grid search or random search. Validate the models on the validation set to monitor performance and prevent overfitting.

### 3.6 Algorithm description:

#### 1. SVM (Support Vector Machine):

Support Vector machine is a supervised machine learning technique. It is used to find a hyper planein a n-dimensional space that distinctly classifies the data points. The hyper plane should have the maximum margin i.e plane having maximum distance between the different classifications of data points.

#### 2. Random Forest Algorithm:

Random Forest is an ensemble learning method that combines multiple decision trees to improve performance and reduce overfittting. Each tree in the forest is trained independently on a random subset of the training data and a random subset of features.

#### 3. Decision Trees:

Decision Trees are hierarchical structures that recursively partition the dataset into smaller subsets based on the values of input features. Each node in the tree represents a decision point based on a feature, and each leaf node represents a class label (spam or legitimate). The tree is built to maximize the information gain or purity at each split.

### 3.7 Techniques:

- **Bootstrap Aggregating (Bagging):** Randomly sample subsets of the training data (with replacement) to train multiple decision trees independently.
- **Hyperparameter Tuning:** Optimize hyperparameters such as the number of trees (n_estimators), maximum depth of each tree (max_depth), minimum number of samples required to split a node (min_samples_split).
- **Tree Pruning:** Implement pruning techniques such as cost-complexity pruning(or post-pruning) to prevent overfitting and improve the generalization ability of the decision tree model.
- **Ensemble Techniques:** Utilize techniques such as stacking or blending to combine predictions from multiple base models effectively.
- **Feature Randomization:** Randomly select a subset of features for consideration at each node of the decision tree within the Random Forest.

## 3. RESULT

It has successfully demonstrated the power of an integrated approach, combining Machine Learning and Deep Learning techniques with both numerical and image datasets to improve the detection and classification of pancreatic cancer.

The machine learning which uses random forest and naive byes yield for a perfect output to identify and classify different pancreatic cancer using the numerical values updated to it based on the update dataset, while deep learning ensures to identify wheather the uploaded CT image has pancreatic cancer or not by scanning the image through it.

The Deep Learning section, adopting the CNN model architecture, excelled with a Training accuracy of 98.7% and Validation accuracy of 100%.This segment emphasized the significance of visual data in healthcare and revealed the system adaptability to different data modalities.

The dataset comprising images for normal and pancreatic tumor classes highlights the versatility of the system in addressing a wide array of medical scenarios.By bringing together these strengths the proposed system enhances diagnostic accuracy, comprehensives, and sensitivity in the assessment of pancreatic cancer.

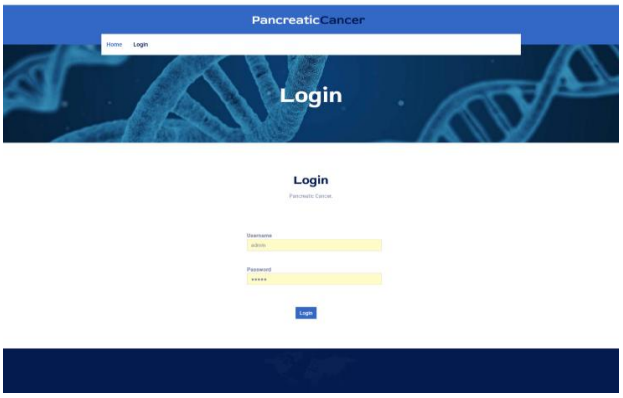


**Figure 4.1: Login Page**

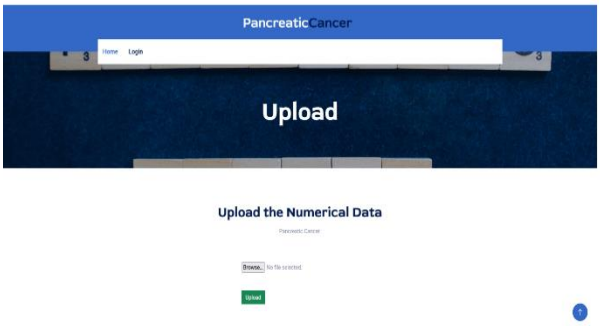


**Figure 4.2: Uploading numerical data for training**

In this we uploading numerical data for training

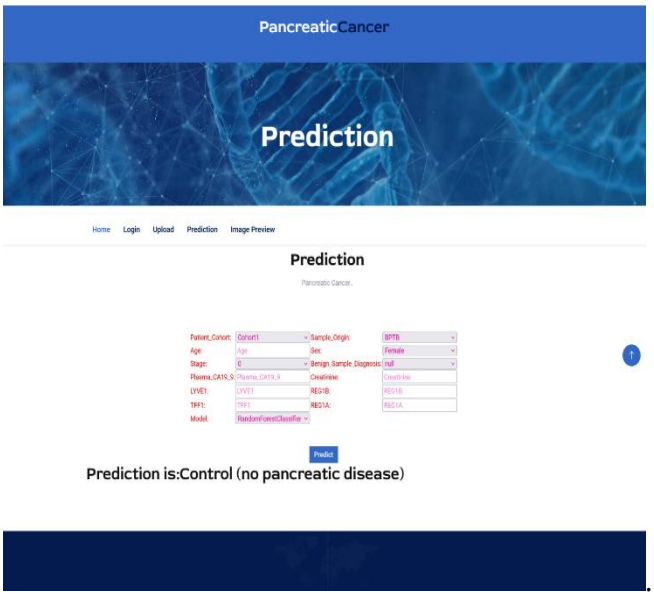


**Figure 4.3: Prediction made based on numerical data**

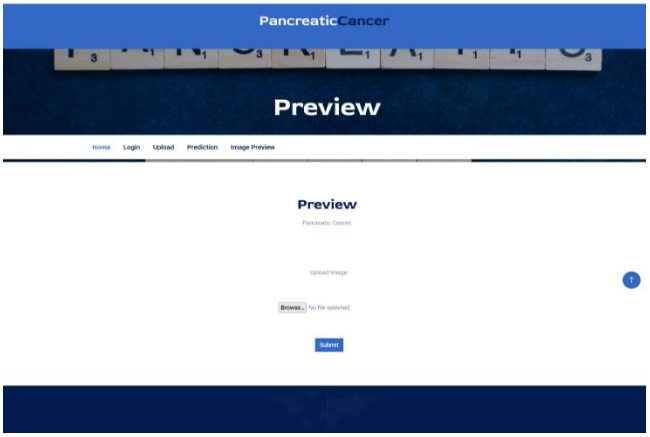


**Figure 4.4: Image uploading page for Prediction**

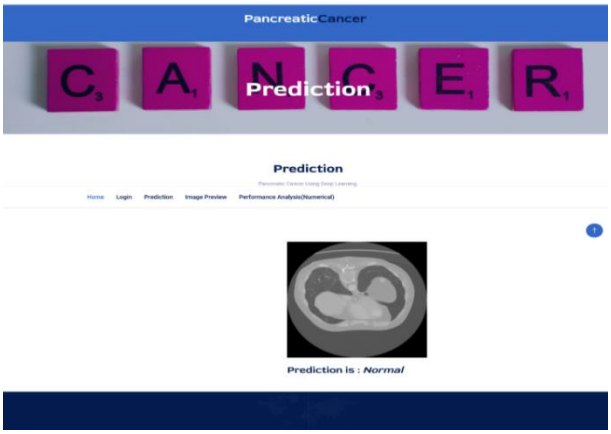


**Figure 4.5: Prediction based on the Image uploaded**

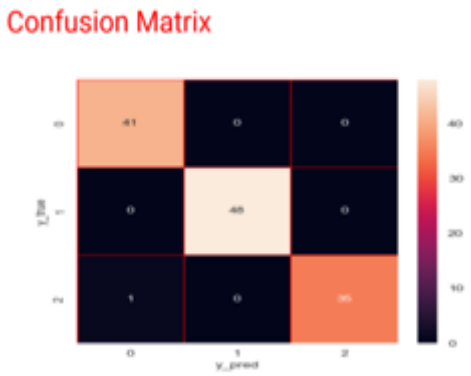


**Figure 4. 6: Confusion Matrix**


**Figure 4.7: Performance Analysis**

## 4. 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 improvingthe 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 ispositive. By using this activation function, the performance of the model is enhanced, and it alsoachieves a higher level of sparsity . We were ableto calculate the likelihood in an effective manner because to the SoftMax layer. When compared tothe previous works, our model achieves outstanding results in terms of its performance, accuracy, and capability to go beyond limits.

## 5. REFERENCE

[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 SupportVector 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] Wismaji Sadewo, Zuherman Rustam, Hamidah Hamidah and Alifah Roudhoh Chusmarsyah; Pancreatic Cancer Early Detection Using Twin Support Vector Machine Based on Kernel.

[5] [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 Patient.

[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.