



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

Prediction of PDAC

Gouri G S¹, Kavya N M², Sachidanand N Kammar³, Sudeep R Shakanavalli⁴, Sanjeev Jigali⁵

UG Student, Department of Computer Science Engineering, Sri Taralabalu Jagadguru Institute of Technology,
Ranebennur, Karnataka, India¹²³⁴

Associate Professor, Department of Computer Science Engineering, Sri Taralabalu Jagadguru Institute of Technology,
Ranebennur, Karnataka, India⁵

ABSTRACT: The human body's pancreas is an organ that can be found in the belly. It is crucial to the process of turning the food we eat into energy for the body's cells. As it is an important procedure in digestion, if the organ gets damaged there will be a problem in the digestion too. As we come to cancer it is a very dangerous disease. When it comes to pancreatic cancer, as the pancreas plays a vital role in digestion, then cancer will affect life severely. Early detection of cancer is very difficult. So, the machine learning technique is used with considering urinary biomarkers as the features of the diagnosis of pancreatic cancer. By using the machine learning technique, we will be able to diagnose cancer at an earlier stage. The algorithm we used is K- Nearest Neighbor, Decision Tree, Random Forest.

I. INTRODUCTION



The human body's pancreas is an organ that can be found in the belly. It is crucial to the process of turning the food we eat into energy for the body's cells. Exocrine function, which aids in digestion, and endocrine function, which controls blood sugar, are the pancreas' two primary tasks. One of the pancreas two primary tasks. One of the most hazardous diseases is pancreatic cancer, according to medical experts. It develops in the pancreatic cells. The pancreas is an organ in your abdomen that is located behind the bottom portion of your stomach.

II. LITERATURE SURVEY

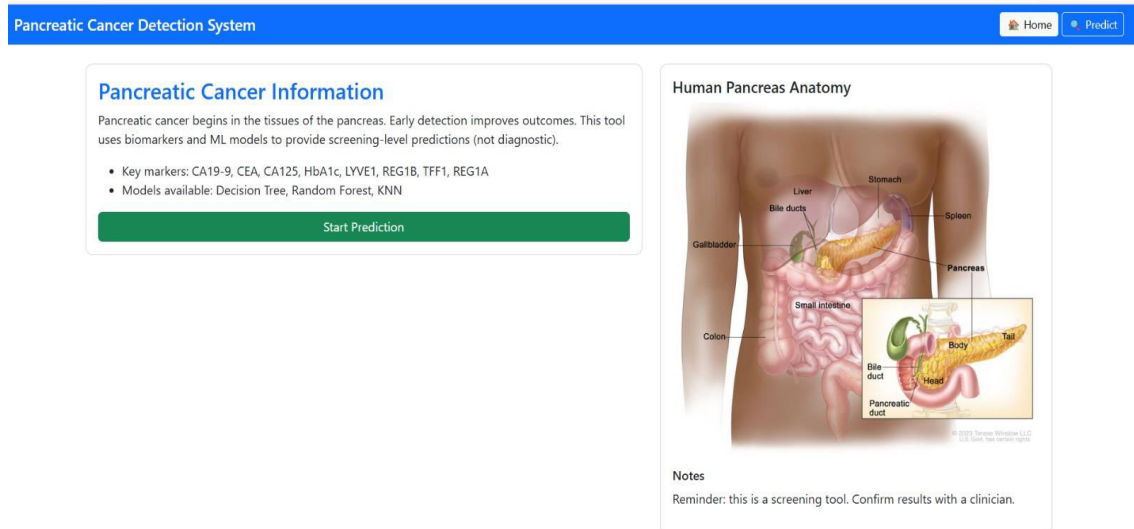
A literature survey is a systematic and comprehensive review of existing research, publications, and scholarly articles related to a specific topic. It involves collecting, analyzing, and summarizing past work to understand the current state of knowledge and identify gaps or areas for further investigation.

Author	Method	Features	Challenges
Silvana Debernardi et al. PLoS Medicine 2020	Urinary biomarker panel (LYVE1, REG1B, TFF1) with PancRISK score algorithm	Non-invasive detection through urine sample High sensitivity and specificity for early-stage PDAC	Limited number of early-stage PDAC sample(n=27)
Sarfaraz Hussein IEEE 2020	Convolutional Neural Network (CNN)	Optimizes spatial redundancy Helps to identify lump automatically	Increases complexity
Wazir Muhammad Frontiers in Artificial Intelligence 2019	Distribution estimation algorithm	Best diagnostic performance Strong ability in classification	High cost. Not fit for huge dataset
Xiao Chen Cancer Management and Research. 2019	Support Vector Machine	High average classification accuracy Minimal error rate	Analysis has to be carried out on large samples
Rong Chen Cancer Management and Research. 2018	Quantitative Computed Tomography Texture Analysis	Reduced False-Negatives	Over-fitting Low precision and sensitivity
Yingfei Tan ISCAI 2017	Large Margin Distribution Machine	High sensitivity Reduced cost	Need to reduce the rate of diagnosis

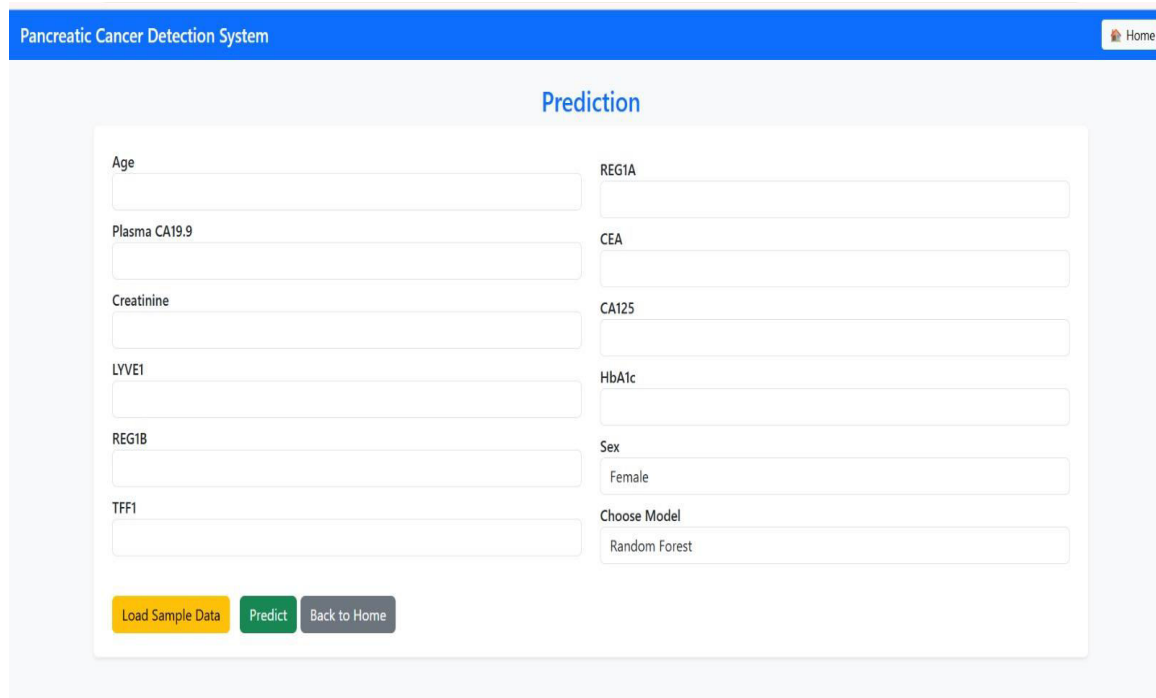
III. METHODOLOGY

The methodology for PDAC prediction involves a systematic process beginning with the collection of high-quality clinical, demographic, imaging, and molecular data from reliable medical datasets and hospital records. The collected data typically includes patient history, blood biomarkers, genetic mutations, imaging findings, and other risk factors associated with pancreatic cancer. After data acquisition, pre-processing is performed to clean missing values, remove errors, normalize numerical features, and encode categorical variables to ensure consistency. Important predictive features are identified using statistical correlation, feature selection techniques, or model-based importance scoring to highlight factors strongly linked to PDAC. The target variable is defined as PDAC-positive or PDAC-negative based on confirmed clinical diagnosis. Machine learning models such as Random Forest, KNN, Decision Tree models for imaging data are trained using a train-test split and cross-validation to ensure robustness. Model performance is evaluated using metrics like accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC-ROC to determine the most reliable predictor. Finally, the best-performing model is used to predict PDAC risk in new patient data, and the results are analyzed to identify key clinical indicators and support early detection research.

IV. RESULT AND DISCUSSION

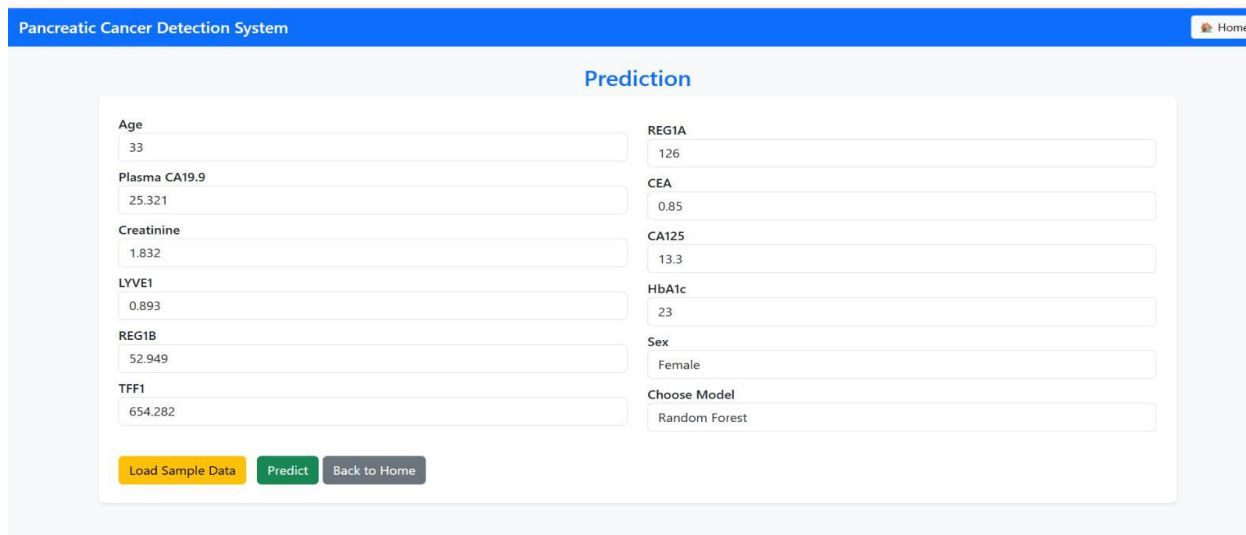


Display a pancreatic cancer screening tool interface, showing cancer information, a “Start Prediction” button, and an anatomy image of the pancreas.



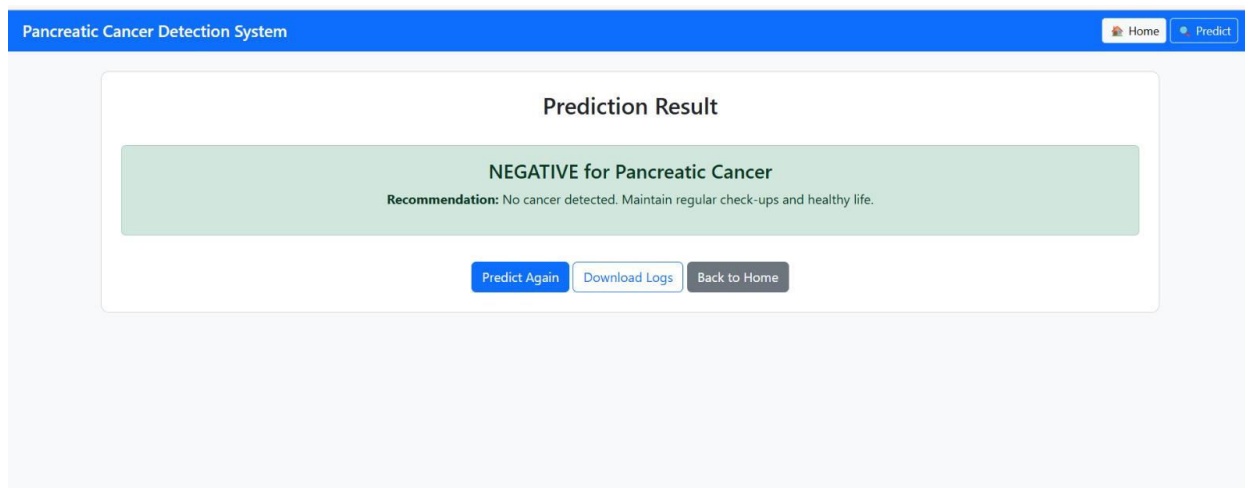
The screenshot displays the 'Pancreatic Cancer Detection System' interface, specifically the 'Prediction' form. The form contains input fields for various medical markers: Age, Plasma CA19.9, Creatinine, LYVE1, REG1B, TFF1, REG1A, CEA, CA125, HbA1c, and Sex. A 'Choose Model' dropdown menu is set to 'Random Forest'. At the bottom, there are three buttons: 'Load Sample Data', 'Predict', and 'Back to Home'.

Displays the picture shows a *Pancreatic Cancer Detection System* web page. It contains a Prediction form with input boxes for various medical markers (like Age, CA19.9, CEA, Creatinine, etc.), a Sex field, a Choose Model dropdown, and buttons for Load Sample Data, Predict, and Back to Home.



The screenshot shows the 'Prediction' form of the 'Pancreatic Cancer Detection System'. The form is divided into two columns of input fields. The left column contains: Age (33), Plasma CA19.9 (25.321), Creatinine (1.832), LYVE1 (0.893), REG1B (52.949), and TFF1 (654.282). The right column contains: REG1A (126), CEA (0.85), CA125 (13.3), HbA1c (23), Sex (Female), and Choose Model (Random Forest). At the bottom of the form are three buttons: 'Load Sample Data' (yellow), 'Predict' (green), and 'Back to Home' (grey). The page header includes the system name and a 'Home' link.

The picture shows a Pancreatic Cancer Detection System webpage with a prediction form where the user enters medical values and selects a model to get a prediction.



The screenshot shows the 'Prediction Result' page of the 'Pancreatic Cancer Detection System'. The main result is 'NEGATIVE for Pancreatic Cancer' displayed in a green box. Below this, a recommendation states: 'Recommendation: No cancer detected. Maintain regular check-ups and healthy life.' At the bottom are three buttons: 'Predict Again' (blue), 'Download Logs' (light blue), and 'Back to Home' (grey). The page header includes the system name and links for 'Home' and 'Predict'.

It shows a webpage displaying a prediction result saying “NEGATIVE for Pancreatic Cancer” with options to predict again, download logs, or go back home.

V. CONCLUSION

With the help of AI/ML code using python, we can analyze any data set given, and as per the topic chosen about the prediction of pancreatic cancer using urinary biomarkers, the dataset extracted here from the kaggle website which contains details about biomarker values of patients with the positive and negative diagnosis. This information and with the help of different algorithms and imputation methods which comes under machine learning and feature engineering helps us to predict the diagnosis of patients with positive and negative pancreatic cancer. In a comparison of all the evaluation parameters of all the algorithms concerning missing, value imputation methods hereby we can conclude that the decision tree algorithm with knn imputation has performed well for the used dataset.

REFERENCES

- [1] Chirath Kumarasiri, Cassim Faroo,” User Centric Mobile Based Decision-Making System Using Natural Language Processing (NLP) and Aspect Based Opinion Mining (ABOM) Techniques for Restaurant Selection”. Springer 2018. DOI: 10.1007/978-3-030- 01174-14
- [2] Atharva Kulkarni,Divya Bhandari,Sachin Bhoite.A study of Restaurants Rating Prediction using Machine Learning Algorithms. International Journal of Computer Applications Technology and Research, 2019, p.377- 378.DOI: 10.7753/IJCATR0809.1008
- [3] Linear Regression, Retrieve from: [HTTPS://WWW.toworddatascience.com/Linear_ Regression. HTM](https://www.toworddatascience.com/Linear_Regression.HTM), LAST Accessed: 5 October 2019
- [4] L.Breiman,Random, Forests.Machine Learning, 45(1), (2001);5–32. <https://doi.org/10.1023/A:1010933404324>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details