WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

AI Driven Pancreatic Cancer Detection

Rithik CA, Shakthivel Parimalam KPC, Venkatraman.

UG Students, Sri Shakthi Institute Of Engineering And Technology, Coimbatore.

ABSTRACT:

Due to the limited options in treatment, prognosis, and symptomless nature, pancreatic cancer is considered to be one of the deadliest types of cancer. Early identification is a must to improve significant survival rates. This can be done by integrating several data sources, such as clinical and imaging data, to propose an AI driven method for the early diagnosis of pancreatic cancer. Cutting-edge methods like convolutional neural networks and machine learning models like random forest classifiers are used in the system to detect early indicators of pancreatic cancer. Doctors and professionals would find this system useful as it allows them to understand the results easily. Explainable AI techniques are employed to make the decision making process better and more transparent. The model's performance will be evaluated using different metrics to ensure effectiveness. This approach can improve patient outcomes and pave the way towards early diagnosis.

Keywords: Pancreatic Cancer, Machine Learning, Convolutional Neural Networks, Random Forests

Introduction:

This paper presents an overview of pancreatic cancer, emphasizing its insidious nature, diagnostic challenges, and contributing risk factors. Often referred to as the "Silent Killer," pancreatic cancer remains one of the most lethal malignancies, accounting for approximately 7% of all cancer-related deaths globally. Despite advancements in medical diagnostics, its detection typically occurs at advanced stages due to the lack of specific early symptoms, resulting in a markedly poor prognosis. Current statistics reveal that only 10% of patients survive beyond five years, and merely 20% of cases are diagnosed at an early, more treatable stage.

The etiology of pancreatic cancer is multifactorial, involving a complex interplay of genetic predispositions and environmental exposures. Key risk factors include age, tobacco use, excessive alcohol consumption, physical inactivity, obesity, chronic pancreatitis, diabetes, family history of pancreatic malignancies, prolonged exposure to environmental pollutants, and dietary imbalances. These variables act independently, yet synergistically contribute to disease progression and severity. Understanding the cumulative impact of these factors is essential for developing targeted screening strategies and preventive interventions.

In the broader context of oncology, pancreatic cancer underscores the critical need for innovative diagnostic tools and public health efforts to mitigate risk. Early detection remains the most effective strategy for improving survival outcomes. As research continues to explore biomarker-based screening and imaging technologies, a multidisciplinary approach integrating lifestyle modification, genetic counseling, and environmental risk assessment is essential. This comprehensive perspective not only highlights the urgency of early intervention but also calls for a unified response in addressing one of the most formidable challenges in modern cancer care.

Risk Factor	Impact on Pancreatic Cancer Development (%)	Notes
Family History	10%	Genetic predisposition increases the risk significantly
Chronic Pancreatitis	5-10%	An inflammatory condition that can lead to malignancy.
Diabetes	5-7%	Prolonged diabetes is linked to a higher cancer risk.
Obesity	3-5%	Increased BMI correlates with cancer incidence.
Genetic Mutations	5-15%	BRCA2, PALB2, and CDKN2A mutations are known contributors.

Spotting pancreatic cancer plays a key role in helping patients live longer. When doctors find it soon enough, they can start treatment that works better. Right now, doctors use pictures to check for pancreatic cancer. But the best pictures we have today don't tell us much about how the disease might go. This leaves doctors with fewer ways to treat it. Taking pictures of the body has become the go-to method to find out if someone has pancreatic cancer, how far it has spread, and what might happen next.[5]. Though there are advancements in the analysis of imaging data and the technologies that have been risen, still there is no guarantee that pancreatic cancer can be detected easily. Not only is this time-consuming, but it also increases the probability of mortality in the affected individual due to the inordinate delay. Nearly 90% of the misdiagnosis of pancreatic cancer was due to the difficulty in spotting the underlying tumour mass due to the inflammation [6]. Thus, there is a pressing need for innovative approaches that enable early detection of pancreatic cancer before symptoms arise. AI-driven methods, with deep learning and machine learning techniques, have great potential in analysing datasets like imaging and clinical datasets and even data with different modalities to detect patterns indicating early-stage cancers. These methods provide a means of identifying key features that can guide early diagnosis and treatment strategies.

Our team's working on a set-up to merge patient health details and picture scans to spot pancreatic cancer sooner. Algorithms like convolutional neural networks will be super important in making a strong setup. This will make the system better at saving lives and giving folks a fighting chance. This study could change the game in spotting pancreatic cancer by giving doctors a way to find it without having to cut people open. It's super precise and quick. Catching the disease fast is going to make a huge difference for patients and might cut down the number of people who die from this super tough sickness.

What is the AI Pancreatic Cancer Detection?

AI Pancreatic Cancer Detection is an app we're building that helps detect pancreatic cancer at an early stage. Users (typically clinicians) input patient health data and medical images, and the app uses artificial intelligence to analyze them for signs of cancer. It identifies patterns in clinical biomarkers and CT scans, then provides an assessment based on medical research and diagnostic models. The app is designed to support doctors with accurate, fast, and non-invasive predictions, improving early detection and patient outcomes.What is the use?

 AI Driven of Pancreatic Cancer Detection is a clinical decision support tool that leverages multimodal data (clinical biomarkers and CT imaging) and advanced machine learning (Random Forest and CNN) to identify early indicators of pancreatic malignancy. It delivers non-invasive, rapid, and accurate analyses by processing urinary biomarker profiles (e.g., LYVE1, REG1B) alongside high-resolution scans, achieving near-perfect classification metrics.

 Explainable AI components such as SHAP value analysis and saliency mapping ensure transparent, interpretable outputs, fostering clinician trust and facilitating risk stratification.

 Clinical Impact: By enabling earlier diagnosis—often challenging due to asymptomatic progression—this system promises to improve patient outcomes, guide personalized treatment planning, and integrate seamlessly into existing workflows for proactive intervention, reducing reliance on invasive biopsy procedures and supporting scalable deployment in clinical settings.

Methodology:

This section outlines the comprehensive methodological framework adopted for the development of an AI-driven pancreatic cancer detection system. The pipeline integrates clinical and imaging datasets, advanced machine learning algorithms, and explainable AI to facilitate early diagnosis and enhance clinical decision-making.

1. Data Collection

- **Objective:** To compile robust and diverse datasets representing both clinical and imaging domains for the detection of pancreatic cancer.
- **Sources:**
    - *Clinical Dataset:* Publicly available dataset from Kaggle containing urinary biomarkers and patient demographics.
    - *Imaging Dataset:* CT scan images representing normal and cancer-affected pancreas, consisting of over 1000 labeled samples.
- **Data Types:**
    - **Clinical Modality:** Age, sex, diagnosis status, cancer stage, benign conditions, and urinary biomarkers (e.g., LYVE1, REG1B, TFF1, creatinine).
    - **Imaging Modality:** High-resolution 2D CT scan slices labeled as "Normal" or "Pancreatic Tumour".

2. Data Preprocessing

- **Clinical Data:**
    - Removal of missing or inconsistent records.
    - Encoding of categorical variables (e.g., sex, diagnosis status).

- Normalization of numerical variables for uniformity.

- **Imaging Data:**
  - Standardization of image size and resolution.
  - Application of augmentation techniques (e.g., flipping, rotation) to enhance training diversity.

3. Feature Engineering

- **Objective:** To extract, analyze, and select the most impactful features for predictive modeling.
- **Techniques Applied:**
  - **Univariate Analysis:** Histograms and frequency plots to assess distributions (e.g., REG1A levels).
  - **Bivariate Analysis:** Box plots and cohort-wise comparisons (e.g., gender-wise diagnosis proportions).
  - **Multivariate Analysis:** Correlation heatmaps to understand relationships among biomarkers.
- **Outcome:** Selection of high-impact predictors such as LYVE1, REG1B, and creatinine for model training.

4. Model Selection and Training

A. Clinical Data Model: Random Forest Classifier (RFC)

- **Reason for Selection:** High interpretability and suitability for structured, high-dimensional data.
- **Workflow:**
  - Train-test split: 80/20.
  - Ensemble learning using multiple decision trees.
  - Majority voting to determine final classification: Control, Benign, or PDAC (Pancreatic Ductal Adenocarcinoma).
- **Performance:** Achieved 100% accuracy on test data with perfect confusion matrix and classification metrics.

B. Imaging Data Model: Convolutional Neural Network (CNN)

- **Reason for Selection:** Superior performance in image recognition tasks.
- **Workflow:**
  - CNN architecture with convolution, pooling, and fully connected layers.
  - Training with annotated CT images, followed by validation on unseen data.
- **Performance:** Achieved 98%–100% accuracy, with no false positives or false negatives.

5. Model Evaluation

- **Metrics Used:**
  - Accuracy, Precision, Recall, F1-Score.
  - Confusion Matrix (perfect diagonal matrix for both models).
- **Results:**
  - RFC and CNN models both achieved **1.00** across all metrics, indicating flawless classification during experimental testing.

6. Integration of Explainable AI (XAI)

- **Clinical Data (RFC):**
  - **SHAP Values** were computed to visualize and quantify each feature's impact on predictions.
- **Imaging Data (CNN):**
  - **Saliency Maps** were generated to highlight regions in CT images that influenced classification, ensuring that AI focused on medically relevant features.
- **Outcome:** Improved transparency, fostering trust in model predictions among clinical practitioners.

7. Deployment Framework and Future Implementation

- **Architecture Integration:** Backend system designed to accommodate real-time inference on clinical and imaging data inputs.
- **Goal:** Seamless integration into clinical workflows and EHR systems for real-world applicability.
- **Planned Enhancements:**
    - Incorporation of genomic and multi-omics data.
    - Expansion to real-time cloud-based deployment.
    - Continuous learning through physician feedback loops.

Results

1. F1 Score

For our clinical dataset where Random Forest Classifier was applied, we achieved an F1 Score of 1.0, conveying optimum performance regarding balance between precision and recall. The other dataset of images, using CNN model applied, also gets F1 Score 1.0; so conceptual performance is at par.

2. Precision

CONCLUSION AND FUTURE SCOPE The Random Forest Classifier for clinical data presented 1.0 as precision, meaning that all predicted instances as positive were accurate. CNN model applied on the imaging dataset also presented precision of 1.0, indicating entirely correct positive predictions of this model. This precision value is quite important, especially in the clinical field, where false positives may lead to unnecessary interventions.

3. Recall

For the Random Forest Classifier of clinical data, the recall value for this model is 1.0 which tells us that the model identifies all true positives. Likewise, the CNN model for the imaging dataset also achieved a recall of 1.0, indicating that every applicable feature was perfectly identified from the images. Since both models showed a recall value of 1.0, it can be concluded that both are extremely suitable in identifying true positive cases; a qualitative aspect in clinical diagnosis and image analysis.

4. Accuracy

The Random Forest Classifier and the CNN model got an accuracy equal to 1.0 on both clinical and imaging datasets, respectively. This means that both models give perfect predictions of all instances in their respective datasets, corresponding to reliable and consistent predictions. Thus, there is not a single instance of misclassification in the prediction of data

Conclusion

This study has made great strides in using two advanced machine learning models: the Random Forest Classifier for clinical datasets and the CNN for medical imaging; both achieving an astounding 1.0 accuracy. One excels in the domain of clinical data, while the other can classify complex medical images. Both models seem suitable for real-world healthcare applications. The methodology established here validates the potential for these models to enhance diagnostic accuracy, aid early diagnosis, and eventually effuse patient outcomes. With the inclusion of rigorous data preparation, model training, and evaluation, the study adds a valid impetus to the already burgeoning efforts to usher machine learning into the field of medical decision-making.

Future Scope

The AI-assisted early detection system for pancreatic cancer corresponds to a massive leap forward among diagnostic methodologies. However, there are several areas where it could welcome innovation for enhanced clinical relevance, accuracy, and effectiveness.

Advancing genomic data would provide the tools for genetic sequencing and biomarker analysis to enhance risk modelling concerning genetic predisposition and the molecular framework underlying pancreatic cancer.

Another development would be the integration of some fusion techniques for multimodal data. It refers to the combination of different data types, such as combining imaging scans, blood test results, clinical history, and genomic data into one diagnostic model. This effort could include augmenting the accuracy of predictive modelling by pooling knowledge from different data sources complementing one another, thus reducing false-positive and false-negative rates.

Besides, just real-time implementing it would enable such a system to be useful to the clinicians. The seamless incorporation into current hospital systems, compatibility with EHRs, and traditional workflows of diagnoses would ease the process of uptake. This could also be made possible through cloud-based AI systems or edge computing techniques to swiftly analyse patient data securely and privately.

References:

List all the material used from various sources for making this project proposal

Research Papers:

1. Rawla, P., Sunkara, T., & Gaduputi, V. (n.d.). *Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors*.
2. Huang, B., Huang, H., Zhang, S., et al. (n.d.). *Artificial Intelligence in Pancreatic Cancer*.
3. Kenner, B., Chari, S. T., Kelsen, D., et al. (2020). *Artificial Intelligence and Early Detection of Pancreatic Cancer: Summative Review*.
4. Hameed, B. S., & Krishnan, U. M. (n.d.). *Artificial Intelligence-Driven Diagnosis of Pancreatic Cancer*.
5. Kamisawa, T., Wood, L. D., Itoi, T., & Takaori, K. (2016). *Pancreatic Cancer*. *The Lancet*, 388, 73–85.
6. Liu, S.-L., Li, S., Guo, Y.-T., Zhou, Y.-P., Zhang, Z.-D., Li, S., & Lu, Y. (2019). *Establishment and Application of an Artificial Intelligence Diagnosis System for Pancreatic Cancer with a Faster Region-Based Convolutional Neural Network*. *Chinese Medical Journal*, 132, 2795–2803.
7. Tripathi, S. (2021). *Artificial Intelligence: A Brief Review*. In *Analyzing Future Applications of AI, Sensors, and Robotics in Society*. IGI Global: Hershey, PA, USA.
8. Rajkomar, A., Dean, J., & Kohane, I. (2019). *Machine Learning in Medicine*. *New England Journal of Medicine*, 380, 1347–1358.
9. Ramesh, A. N., Kambhampati, C., Monson, J. R. T., & Drew, P. J. (2004). *Artificial Intelligence in Medicine*. *Annals of The Royal College of Surgeons of England*, 86, 334–338.
10. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). *AI in Health and Medicine*. *Nature Medicine*, 28, 31–38.
11. Zhang, Y., Wang, S., Qu, S., & Zhang, H. (2020). *Support Vector Machine Combined with Magnetic Resonance Imaging for Accurate Diagnosis of Paediatric Pancreatic Cancer*. *IET Image Processing*, 14, 1233–1239.
12. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (n.d.). *The Consensus Molecular Subtypes of Colorectal Cancer*.
13. Jonckheere, N., Auwercx, J., Hadj Bachir, E., Coppin, L., Boukrout, N., Vincent, A., et al. (2020). *Unsupervised Hierarchical Clustering of Pancreatic Adenocarcinoma Dataset from TCGA Defines a Mucin Expression Profile That Impacts Overall Survival*. *Cancers*, 12, 3309.
14. Moffitt, R. A., Marayati, R., Flate, E. L., Volmar, K. E., Loeza, S. G. H., Hoadley, K. A., et al. (2015). *Virtual Microdissection Identifies Distinct Tumor- and Stroma-Specific Subtypes of Pancreatic Ductal Adenocarcinoma*. *Nature Genetics*, 47, 1168–1178.
15. Bohr, A., & Memarzadeh, K. (2020). *Artificial Intelligence in Healthcare*. Academic Press: Cambridge, MA, USA. ISBN 9780128184394.
16. Abel, L., Wasserthal, J., Weikert, T., Sauter, A. W., Nesic, I., Obradovic, M., et al. (2021). *Automated Detection of Pancreatic Cystic Lesions on CT Using Deep Learning*. *Diagnostics*, 11, 901.
17. Dinesh, M. G., Bacanin, N., Askar, S. S., & Abouhawwash, M. (2023). *Diagnostic Ability of Deep Learning in Detection of Pancreatic Tumour*. *Scientific Reports*, 13, 9725.
18. Viriyasaranon, T., Chun, J. W., Koh, Y. H., Cho, J. H., Jung, M. K., Kim, S.-H., et al. (2023). *Annotation-Efficient Deep Learning Model for Pancreatic Cancer Diagnosis and Classification Using CT Images: A Retrospective Diagnostic Study*. *Cancers*, 15, 3392.
19. Huang, J., Zhou, Y., Zhang, H., & Wu, Y. (2023). *A Neural Network Model to Screen Feature Genes for Pancreatic Cancer*. *BMC Bioinformatics*, 24, 193.
20. Chen, P.-T., Wu, T., Wang, P., Chang, D., Liu, K.-L., Wu, M.-S., et al. (2023). *Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-Based Study*. *Radiology*, 306, 172–182.