IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Diagnosis And Detection Of Pancreatic Cancer Using Advanced Learning Models And Explainable Artificial Intelligence

Meghna Panara, Vikas Tulshyan

[1]Research Scholar, Silver Oak University, Ahmedabad, Gujarat, India

[2] Associate Professor, Silver Oak University, Ahmedabad, Gujarat, India

Abstract

Pancreatic cancer (PC) is still one of the deadliest types of cancer. It is the 10th most prevalent malignancy in males and the 7th most frequent among females. The aim of this review paper is to thoroughly analyze the research conducted by researchers on the detection of pancreatic cancer by employing several artificial intelligence learning methods such as deep learning, machine learning etc. The PRISMA approach was employed for conducting a comprehensive assessment to include relevant papers published between the years 2019 and 2024. This study emphasizes the significance of current methods in identifying and diagnosing pancreatic cancer. The compilation of data from research studies provided an outline of the present state of AI applications in this field. It emphasized the advances that had been made and the ongoing challenges that must be addressed to enhance diagnostic accuracy and patient outcomes. While examining the existing techniques, it had been found that the perfect scores had been computed by YOLOCNN in terms of accuracy, precision, recall, and F1 score while as a neuro-fuzzy interference system achieved 99.95% accuracy. Additionally, 3D UNet attained 100% sensitivity and specificity with an AUC-ROC of 0.99. The novelty of the study lies in its integration of advanced convolutional neural networks with the explainable artificial intelligence to enhance model interpretability and trustworthiness which thereby facilitates better clinical decision-making and understanding of the diagnostic process for pancreatic cancer.

Keywords: Pancreatic cancer, Machine learning, XAI, Deep learning, Advanced CNN, Feature extraction

1. Introduction

Pancreatic cancer ranks as the fourth most significant contributor to cancer-related mortality in both males and females. Due to its hostile nature and delay in diagnosis, it remains one of the most challenging forms of cancer to treat which leads to drastic reduction in the survival rate. It is less common than cancers like breast, lung, as well as colorectal cancer but its diagnosis is challenging. Albeit, age is a main risk factor, but it is impractical to screen the entire population on the basis of age as it includes high cost and risk of false positives [1]. Few high-penetrance risk factors are also known such as family history, genetic mutations, hereditary conditions, chronic pancreatitis, smoking etc which hinders its early detection. In 2024, the American Cancer Society displays that around 66,440 people in the States will be detected with pancreatic cancer, with slightly more cases among men (34,530) than women (31,910). Tragically, about 51,750 people are expected to succumb to the disease during the same period, with a higher mortality rate among men (27,270) compared to women (24,480) [2]. Hence, it necessities continued research efforts into early detection methods and more effective treatments of pancreatic cancer. Figure 1 defines the annual increase of different cancers from 2019 to 2024 [3].

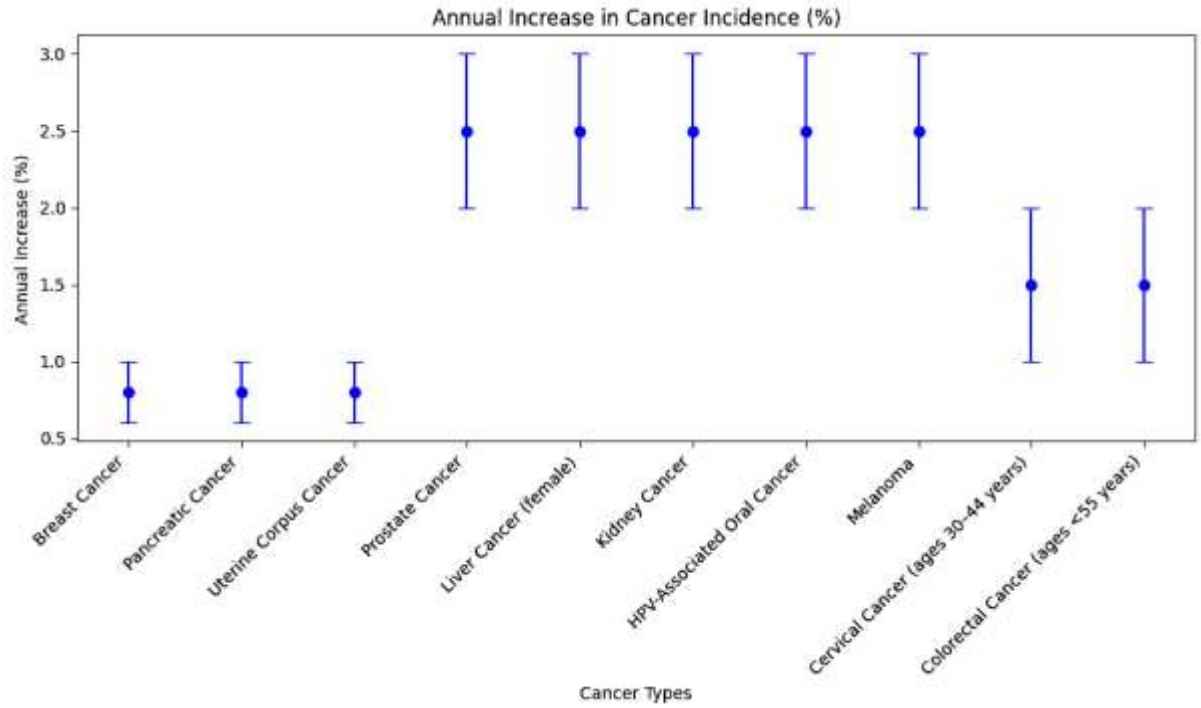

Figure 1: Annual increase of multiple cancers

There are **traditional techniques** that can detect as well as diagnose pancreatic cancer. These techniques involve imaging methods, comprehensive medical assessment, laboratory tests, and sometimes surgical interventions. Initially, symptoms such as weight loss, jaundice, abdominal pain, etc are identified via thorough medical history and physical examination to indicate the presence or absence of pancreatic cancer [4]. Imaging techniques play an important role in the diagnosis like magnetic resonance imaging (MRI), computed tomography (CT) scans, and endoscopic ultrasound (EUS), are helpful in detecting tumors, assess their size, and determine if the cancer has spread. Laboratory tests such as blood tests for tumor markers like CA 19-9, can support the diagnosis, although they are not definitive on their own [5]. In some cases, a biopsy is required in which a sample of a tissue is obtained and examined using microscope to confirm the diagnosis [6]. Although, these traditional methods are useful to early detection of pancreatic cancer but they also have few **limitations**. Physical examinations and medical analysis sometimes fail to identify pancreatic cancer at its early stage due to ambiguity in symptoms. In addition to this, imaging techniques may not always detect small tumors or accurately differentiate between benign and malignant lesions [7]. Laboratory tests can also generate false positives and biopsies which are important for confirming the presence of cancer are sometimes also painful as well as risky. These limitations necessitate the need for more accurate, sensitive, and non-surgical diagnostic methods [8].

Incorporating traditional techniques with **learning models** can noticeably enhance the rate of detection, classification, and diagnosis of pancreatic cancer. These algorithms can assess large amount of imaging data for identifying patterns and features that may be unnoticeable to the human eye [9]. Moreover, combining imaging data with clinical and laboratory data, the AI learning models can provide more total risk evaluations as well as personalized diagnostic insights. Apart from this, they can also classify pancreatic lesions more accurately as well as differentiate between benign and malignant tumors with high precision [10]. Hence, combining traditional techniques with AI techniques not only amplifies the diagnostic capabilities but also simplifies early detection, accurate classification, and personalized treatment planning which lead to better patient outcomes in pancreatic cancer care.

1.1 Research Contribution

The main contributions of the study are as following:

1. The study provides brief information on pancreatic cancer, its health statistics, conventional treatment methods, their limitations, and how AI are capable to address these shortcomings.

2. A systematic review of relevant research papers has been conducted using the PRISMA criteria which lead to the formulation of several research questions.

3. The study includes a detailed investigation and analysis of researchers' contributions to the application of machine and deep learning techniques in detecting and classifying pancreatic cancer along with the challenges they encounter.

4. Additionally, the study introduces a proposed methodology involving optimized hybrid deep learning techniques integrated with ExplainableAI (XAI) for the detection and classification of pancreatic cancer.

1.2 Research Questions

Following are the research questions that have been framed and discussed in Section 4

RQ1: How do deep learning-based approaches for pancreatic cancer detection are better than traditional machine learning methods and radiologist interpretations?

RQ2: How can advance learning techniques be used to extract and select the most relevant features effectively from multimodal data sources for the accurate prediction of pancreatic cancer?

RQ3: What are the advantages of Collaborative studies and accessible data programs in advancing deep learning approaches for pancreatic cancer detection?

1.3 Layout

Section 1 provides a comprehensive overview of pancreatic cancer, including both traditional and AI-based diagnostic tools. Section 2 outlines the methodology employed in this study, with a focus on the PRISMA aspect of the article. Section 3 provides information about the research conducted in the same field, followed by Section 4 and Section 5 where the article is discussed and summarized, respectively.

2. Review methodology

This review has compiled with respect to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards which consists of four phases i.e. Identification, Screening, Eligibility, and Included, as shown in Figure 2.

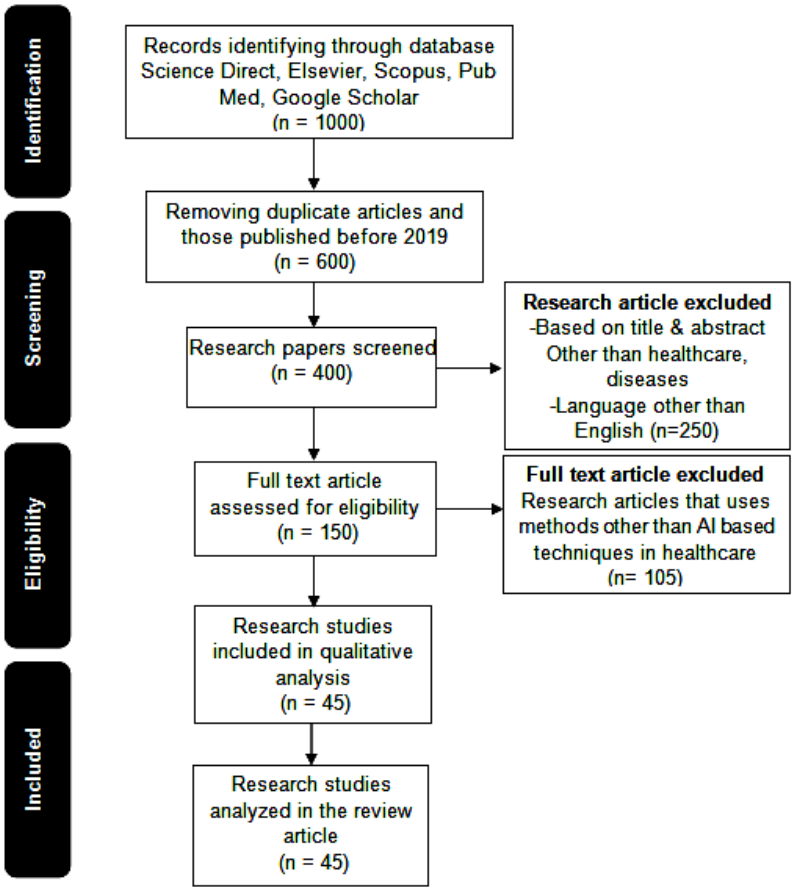

Figure 2 : PRISMA guideline for the selection of papers

In the phase of identification, papers are searched from multiple repositories such as Google Scholar (https://scholar.google.co.in), ScienceDirect (https://www.sciencedirect.com), PubMed (http://www.ncbi.nlm.nih.gov/pubmed), Springer (https://www.springer.com/in), and Scopus (https://www.scopus.com) using the keywords "pancreatic cancer", "machine learning", "artificial Intelligence", "radiologists", "eXplainable artificial intelligence", "traditional techniques", and "deep learning", as well as numerous keyword combinations. Later, in case of screening, transparent selection of papers is performed on the basis of the year, title, abstract etc followed by eligibility phase where insufficient information based papers are excluded. At the end, there is included phase where the filtered out papers are selected and used for further analysis.

3. Literary research

This section provides a detailed review of related work which encompasses studies related to the pancreatic cancer detection and classification. Moreover, the section also addresses key findings and limitations found in existing literature in order to lay the foundation for the contribution of the study to the field, as shown in Table 1.

3.1 Background

Hameed et al. (2022) [11] investigated the potential of AI models to enhance the detection of pancreatic cancer by using various imaging techniques. It also explored the current advancements in AI-based diagnosis by focusing on cytopathology and serological markers. Discussions had being done to address the ethical concerns surrounding the usage of these instruments. Whereas **Viriyasaranon et al. (2023)** [12] developed an innovative deep learning model that could identify pancreatic cancer without the need for extensive training datasets with annotations. The study used CT images obtained between 2004 and 2019 from a cohort of 4287 individuals who were diagnosed with cancer. It was a retrospective diagnostic study where the researchers introduced a self-supervised learning approach called pseudo-lesion segmentation for classification of

pancreatic cancer. The algorithm was trained and verified using randomly divided training and validation sets, both with and without pseudo-lesion segmentation. In addition, they conducted cross-racial external validation by using open-access CT data from 361 patients, in which the models demonstrated strong performance.

Acer et al. (2022) [13] presented a highly encouraging approach for identifying pancreatic ductal adenocarcinoma (PDAC) in its first phase by employing non-invasive urine biomarkers and carbohydrate antigen 19-9 (CA19-9). The researchers worked on the Kaggle Urinary Biomarkers for Pancreatic Cancer (2020) dataset, which is openly accessible and consisted of 590 participants. Multiple machine learning classifiers such as, naive Bayes, support vector machine, (kNN), random forest, k-nearest neighbors, AdaBoost, light gradient boosting machine, and gradient boosting classifier (GBC), were employed in the study. For both binary as well as multiclass classification, the execution and validation were performed using 5-10-fold cross-validation for identifying the optimal machine learning model which can differentiate between individuals with pancreatic problems, healthy controls, and patients with PDAC. **Muhammad et al. (2019)** [14] discussed about the difficulty of recognizing pancreatic cancer in its early stages. The reason behind it was the presence of symptoms that often appear when the cancer has already progressed, as well as the absence of a dependable screening method to detect it. Therefore, they developed an artificial neural network and trained as well as tested it with the data of 800,114 participants obtained from the National Health Interview Survey and 898 patients from Pancreatic, Lung, Colorectal, and Ovarian cancer (PLCO) datasets who had been diagnosed with pancreatic cancer. The neural network was used to incorporate 18 features and assess the risk of pancreatic cancer at an individual level. **Jan et al. (2023)** [15] explored AI models to predict and diagnose pancreatic cancers. The authors included 30 studies from 1185 publications in a scoping review. They stated that most of the papers used AI for diagnosing pancreatic cancer with most frequently used radiological image data. 37% of the studies used datasets having less than 1000 samples. 60% of the papers mentioned about the Deep learning such as CNN for pancreatic cancer diagnosis and only 33% used validation methods such as k-fold cross-validation and external validation for it. The authors mentioned that from the surveyed papers, decision trees, support vector machines, as well as k-means clustering reported the highest accuracy of 99%. **Dinesh et al. (2023)** [16] aimed to assess the efficacy of the innovative YCNN strategy in comparison to other contemporary methods for predicting pancreatic cancer. This study used a Convolutional Neural Network model to accurately CT scan forecast images of pancreatic cancer. Furthermore, the authors employed the YOLO-CNN to assist in the categorizing procedure. Both CT image datasets and biomarkers are utilized for testing. The YCNN approach demonstrated exceptional performance with perfect accuracy in a comprehensive analysis of comparable results, surpassing other contemporary techniques. In their recent study, **Huang and colleagues (2022)** [17] dug into the domain of artificial intelligence and its essential role to upend pancreatic cancer care. Their research focused on using AI algorithms which rapidly identifies high-risk groups within the patient. These algorithms not only predict the likelihood of return of cancer but also estimate life expectancy, spread of cancer, and response to treatment which are critical factors that clearly control the prognosis of patient. N**asief et al. (2019)** [18] worked on the data of daily non-contrast CT scans from 90 patients with pancreatic cancer who were having CT-guided chemoradiation therapy. 2520 CT sets were examined which yielded more than 1300 radiomic characteristics. Strongly correlated dependent variable factors were removed using Spearman correlations, and linear regression models were used to determine the connection between the chosen delta radiomic features and pathological response. Statistically significant changes in delta radiomic features were detected through the use of T tests and linear mixed-effects models while as a response prediction model were constructed using a Bayesian regularization neural network. **Luo et al. (2020)**[19] used deep learning model to analyze the contrast-enhanced computed tomography (CECT) images. The approach was applied to create three distinct models: venous, arterial, and arterial/venous. The CECT pictures were analyzed at their ideal phase to compare the DL models with traditional machine learning models in their ability to predict the abnormal grading of pancreatic neuroendocrine neoplasms. The evaluation also included an assessment of the performance of radiologists based on quantitative as well as qualitative CT findings. **Lee et al. (2021)** [20] employed artificial intelligence methods as well as multi-center registry data for the analysis of surgical lapse of pancreatic cancer and its key factors. Korea Tumor Registry System (KOTUS) was used which had the data of 4846 patients. The Cox proportional-hazards model and random forest algorithm were administered and assessed for predicting disease-free survival.

Ramaekers et al. (2024) [21] developed a framework based on deep learning model to detect pancreatic head cancer accurately using the CT scans incorporated by clinically relevant data. By considering secondary signs of pancreatic tumors and CT scan images, the framework dramatically increased detection accuracy for pancreatic head tumors and showcased that incorporating secondary signs with CT scans escalated the

accuracy of detecting pancreatic tumors. **Shi et al. (2023)** [22] applied machine learning techniques and employed minimally invasive liquid biopsies to enhance the early detection of pancreatic cancer. The researchers collected data from nine publicly available pancreatic cancer dataset. The collection comprised of small RNA sequencing data for non-tumor and tumor samples from patients, as well as serum samples. Systematic analyses were conducted for evaluating serum samples and corresponding tissues. The Robust Rank Aggregation technique was applied to find feature markers that exhibited co-expression in both sample types. Subsequently, the clinical importance of these markers was confirmed in the serum of patients. The efficacy of pancreatic cancer prognosis was analyzed using four machine learning approaches, employing the most promising candidate indicators. Additionally, these markers were assessed for their ability to differentiate between pancreatic cancer and pancreatitis. **Dodda & Muneeswari** (2024) [23] recommended an Improved Adaptive Neuro-Fuzzy Inference System (IANFIS) for classifying pancreatic cancer. Their main objective was to analyse the affected region of tumor in the pancreas images for which Bayesian Hyper parameter optimizer was used to fine tune the parameters of the IANFIS classifier. In addition to this, for reducing the noisy signals in the images, Gabor filtering technique was used during pre-processing phase. Subsequently, Artificial Gorilla Troops Optimizer was used for optimal feature selection followed by segmentation of pancreas as well as the affected tumor portion using Enhanced Red Fox Optimization Algorithm. While execution, it had been found that their proposed optimized machine learning-based classifier increased the accuracy of pancreatic cancer diagnosis. **Ravi et al. (2023)** [24] studied a novel strategy for early diagnosis of pancreatic cancer by using advanced deep learning techniques i.e. combination of Support Vector Machine and Multi-Layer Perceptron. This innovative approach demonstrated positive outcomes to accurately identify pancreatic cancer at its early stages. Multilayer perceptron was used for extracting complicated features from medical data and for classifier support vector machine was employed which achieved exceptional results in terms of accuracy, sensitivity, along with specificity. Their research underline the capability of models for diagnosing pancreatic cancer earlier and highlighted the impact of innovative machine learning technologies in medical sector. **Placido et al. (2023)** [25] stated the need for early detection of pancreatic cancer. They collected the medical data from Danish National Patient Registry, Denmark which included 24000 and 3900 cases of pancreatic cancer from US Veterans Affairs respectively. CancerRiskNet was proposed and trained on the hierarchy of disease codes in health records to predict the occurrence of cancer within sequential time segments. The results obtained by the model were fine and highlighted to work on the ability of design realistic surveillance programs for patients at greater risk to boost their quality of life and lifespan on detecting pancreatic cancer at its early stage. **Tonozuka et al. (2020)** [26] designed an original computer-assisted diagnosis system using deep learning to examine EUS images (EUS-CAD). The capability of the system in detecting pancreatic ductal carcinoma (PDAC) was appraised against images from those with a normal pancreas and patients with chronic pancreatitis. A whole of 920 endosonographic images were considered for training and extra 470 images used for independent testing. They analyzed the detection performance of the system in both validation and test settings as well as identified factors independently related to misdetection of the cancer.

Table 1: Evaluation of the prior research

Authors	Dataset	Techniques	Outcomes	Limitations
Viriyasaran on et al. (2023) [12]	CT images 4287 patients	CNN	Acc = 94.3% Sens = 92.5%	Limited dataset, lack of generalization, discrepancies in validation
		Transformer based DL method	Acc = 95.7% Sens = 99.3%	
Acer et al. (2023) [13]	Healthy controls	GBC	Acc = 92.99% AUC = 0.9761 Recall = 0.9245 Prec = 0.9368 Kappa = 0.8598	Limited Dataset for Urinary Biomarkers, Limited Applicability to Other Types
	Pancreatic disorder	LightGBM	Acc = 86.37% AUC = 0.9348 Recall = 0.8451 Prec = 0.8759 Kappa = 0.7261	

Muhammad et al. (2019) [14]	National Health Interview Survey	Artificial neural network (ANN)	Sens = 87.3% AUC = 0.86 Spec = 80.8%	Low prediction rate and Limited dataset
	PLCO dataset		Sens = 80.7% AUC = 0.85 Spec = 80.7%	
Dinesh et al. (2023) [16]	CT scans for PC	YOLO based CNN	Prec = 100% Recall =100% F1 score =100% Acc = 100%	Lack of interpretability and specific details of the dataset
Hussein et al. (2019) [45]	171 MRI scans of pancreatic tumor	3D CNN with multitask learning	Acc = 91.26% Mean score difference = 0.459	Lack of information, Improvement in the performance of model
Althobaiti, M. et al. (2023) [44]	CT Images	Optimal deep learning-based model	Sens = 97.88% Spec = 99.38% Acc = 98.08% F1 = 98.63%	Unspecified dataset details in terms of size or characteristics of CT image
Shi et al. (2023) [22]	26 pancreatic cancer patients	Robust Rank Aggregation Algorithm	Acc = 91.5%	Low enrollment and single centre study
Nasief et al. (2019) [18]	2520 images of CT scan	Bayesian model	AUC =0.94	Technological modifications and advanced models needs to be required for the cancer detection
Lee et al. (2021) [20]	Real data on the population of South Korea	Random forest classifier	C-index = 0.68	Validation of the research is required in medical practice
		Cox model	C-index = 0.77	
Dodda & Muneeswari (2024) [23]	Publically available pancreatic cancer data	IANFIS	Acc = 99.95%, Sens = 99.87% Spec = 99.92%	-
Luo et al. (2020) [19]	CECT images collected from 93 and 19 patients from hospital I and II respectively	Convolution neural network	Arterial (AUC =0.81)	Reliance on manual localization, misclassification
			Venous (AUC = 0.57)	
			Arterial/Venous (AUC = 0.70)	
			Acc = 88.1% AUC = 0.81	
Ravi et al. (2023) [24]	Pancreatic cancer images	Multi-Layer Perceptron and Support Vector Machine	Accuracy = 98.41%	Optimization is required to enhance the performance of models
Ramachandra et al. (2023) [1]	Samples of Urine	RIC-GD	Accuracy = 92%	The accuracy of the model needs to be improved

Tonozuka et al. (2020) [26]	920 endosonographic images	EUS-CAD system	Sens = 92.4% Spec = 94.1% PPV = 86.8% NPV = 90.7% AUC = 0.924 to 0.940	Lack of sample size calculation, Insufficient number of patients
Placido et al. (2023) [25]	Danish National Patient Registry (DNPR)	CancerRiskNet	AUROC = 0.88	Availability of limited computational resources
	US Veterans Affairs (US-VA)		AUROC = 0.78	
Ramaekers et al. (2024) [21]	59 CT scans	3D U-Net	Sens = 0.97 Spec= 1.00 Acc = 0.98 Prec = 1.00 AUC = 0.99 F1 = 0.98 Mean Dice = 0.34	Small dataset, Model could only discriminate pancreatic head cancer
	28 pancreatic cancer cases		Sens = 1.00 AUROC=0.99 Mean Dice = 0.37	

Sens = Sensitivity, PPV = Positive Predicted Value, NPV = Negative Predicted Value, Spec = Specificity, Acc= Accuracy, Prec = Precision

3.2 Research gap

Based on the limitations identified across various studies on pancreatic cancer detection and diagnosis, several significant research gaps emerge. Many studies suffer from a limited dataset, which hinders the generalization and applicability of their findings. For instance, few researchers noted limitations due to the size and specificity of their datasets, affecting the robustness of their models. Additionally, several studies, lacked adequate sample size calculations or enrolled insufficient numbers of patients, potentially compromising the reliability of their conclusions. Interpretability issues were prevalent in studies, where details about dataset specifics and model operations were insufficiently detailed. Furthermore, disparities between internal and external validation data were noted in few works which highlighted the need for improved validation methodologies. Addressing these gaps through larger, more diverse datasets, enhanced model interpretability, and rigorous validation techniques will be crucial for advancing the accuracy and reliability of pancreatic cancer detection models in clinical practice.

Hence, the proposed methodology integrates advanced Convolutional Neural Networks with explainable artificial intelligence which analyses and classifies the complicated features in pancreatic cancer images. This ability not only advances the accuracy of diagnosis but also provides detailed explanations for making clinical decisions. In fact, addressing such limitations maximizes the use of optimized hybrdizied CNN architectures which can obtain higher accuracy in identifying pancreatic cancer.

4. Discussion

RQ1: How do deep learning-based approaches for pancreatic cancer detection are better than traditional machine learning methods and radiologist interpretations?

Deep learning-based approaches, exemplified by convolutional neural networks (CNNs), differ significantly from traditional machine learning methods in several key aspects. Traditional machine learning techniques like SVMs, logistic regression, random forests etc lean on engineered features that are manually extracted with the help of experts from different type of clinical data [27]. While as CNNs independently or automatically learn hierarchical representations of features which have been taken directly from raw data, as shown in Figure 3. In addition to this, it also eliminates the manual work of extracting features and potentially captures the difficult or complicated patterns that may sometimes have been ignored or missed by traditional approaches. This characteristic of deep learning models often generates superior performance as well as accuracy while dealing with high-dimensional and complex datasets like analysis of medical images [28]. Moreover, advanced deep learning models such as transfer learning provide high level of adaptability as well as scalability that can minimize re-engineering and enhance performance in applications such as pancreatic cancer detection.

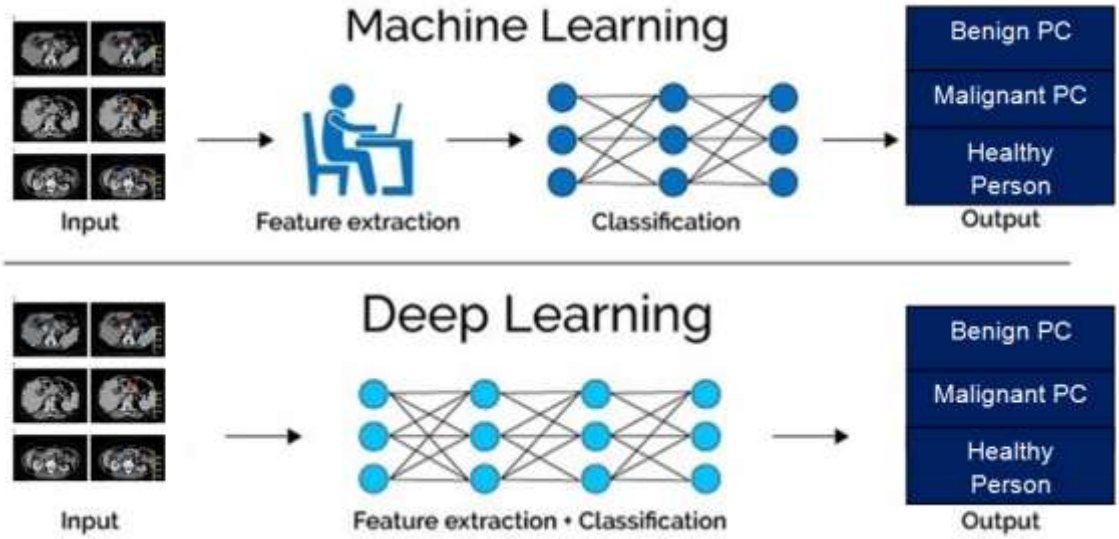

Figure 3 : Machine Learning Vs Deep Learning

In case of radiologist interpretations, deep learning approaches provide consistent and objective assessments as well as are capable to handle large set of data. On the other hand, radiologists provide only subjective interpretations on the basis of their expert knowledge which carries a high chance of variation due to certain factors like workload or fatigue or manual error etc. This problem can badly affect or misinterpret the diagnosis, prediction, and classification of any disease such as pancreatic cancer [29]. As already mentioned the ability of deep learning models on studying the huge bulk of data, they are trained with distinct dataset which either can match or exceed the performance of human for the same task. It happens because the neurons present in the layers of deep neural network are capable to identify subtle or intricate patterns which enhance the accuracy of predicting pancreatic cancer at its earliest stage, as shown in Figure 4 [30]. It has been also noted that radiologists merge the findings obtained from the image with the history of the patient for comprehensive or whole patient treatment but conversely neural network analyse these interpretations by providing statistical analysis which potentially improves the efficiency and diagnostic accuracy in clinical workflows.

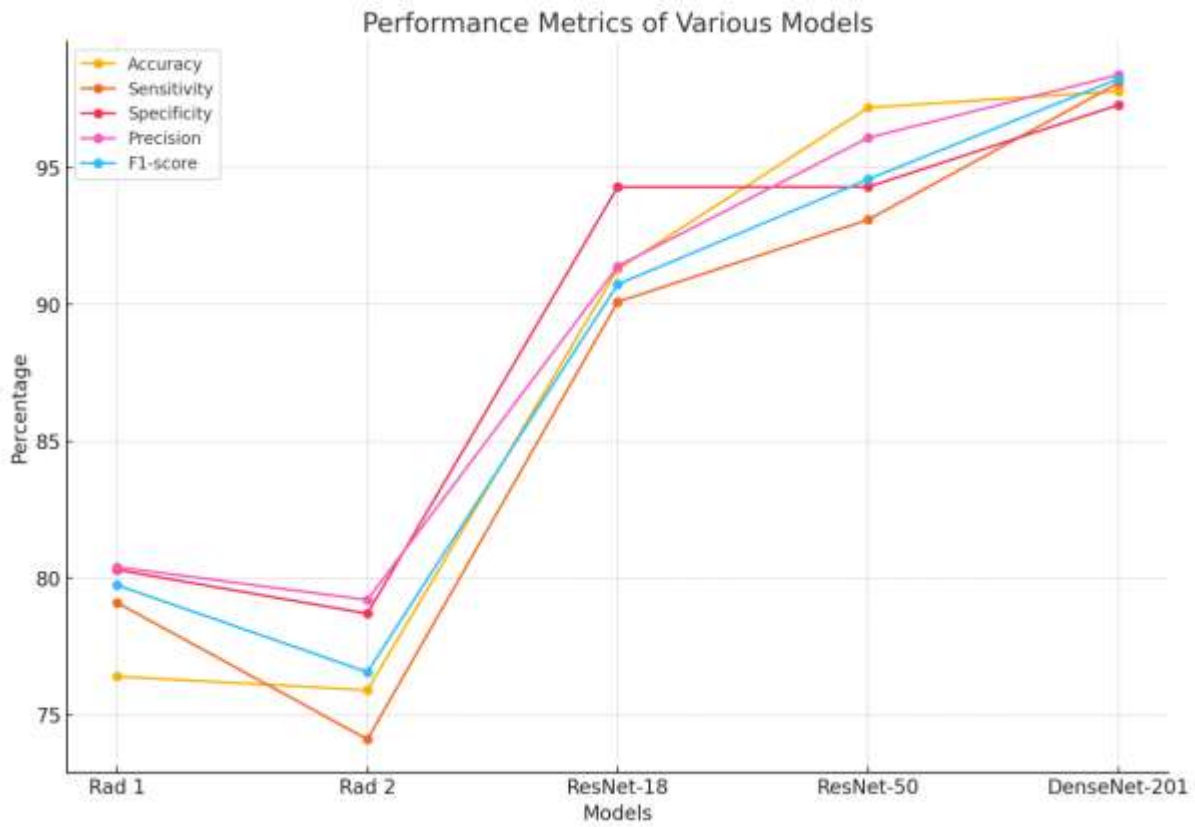

Figure 4 : Deep Learning vs Radiologists

However, there are challenges like the need for extensive datasets, validation across diverse populations, and surety of robustness and interpretability in model must be addressed to fully integrate deep learning into medical practice.

RQ2: How can advance learning techniques be used to extract and select the most relevant features effectively from multimodal data sources for the accurate prediction of pancreatic cancer?

Advanced learning techniques are useful in extracting as well as selecting the most relevant features from multiple data sources to accurately identify pancreatic cancer as shown in Figure 5:

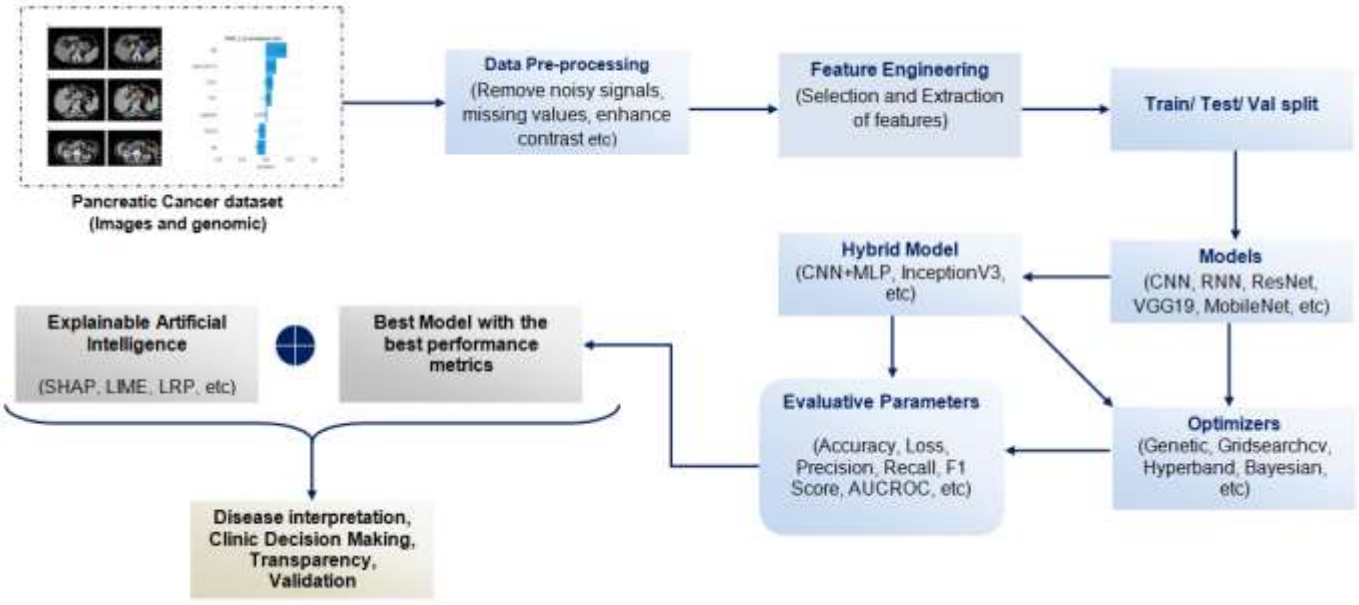

Figure 5 : Proposed design to predict and classify pancreatic cancer

Data Acquisition: In this phase, the data such as genomic data, normalized data, imaging data, and clinical records which are collected from multiple sources are integrated and combined into a set of single feature before feeding it into the model. It ensures that all the relevant information is presented together which enables the model to learn from the dataset effectively.

Data pre-processing: After collecting the different forms data, it is important to pre-process them as image data can vary in resolutions with different lightning conditions and have noisy signals. Hence, to standardize such images, various techniques like Gaussian smoothing (eq i-v), equalization of histogram and resizing can be used.

Compute histogram $H(i)$ where i refers to the intensity values (i)

Normalization of histogram

$$p(i) = \frac{H(i)}{N\ x\ M} \quad \text{(ii)}$$

Here, N and M are dimensions of image

Compute CDF (Cumulative Distributive Function)

$$CDF(i) = \sum_{j=0}^{i} p(j) \quad \text{(iii)}$$

Define Transform Function

$$T(i) = \mu + \sigma.\Phi^{-1}\big(CDF(i)\big) \quad \text{(iv)}$$

Here μ = mean and σ= standard deviation

Apply Transformation

$$I_{equalized}\ (x, y) = T(I(x, y)) \quad \text{(v)}$$

Here, I = input image, x and y= spatial coordinates

Apart from this, it is important to normalize the pixel values also in order to assure its consistency across different image datasets. Moving to genomic or numerical data, there is possibility of having outliers or missing values which can be filled using imputation techniques and normalized using z-score normalization (eq vi) to ensure consistency of data. These pre-processing methods make sure that the data is in a suitable format to analyse and train model which overall enhances their reliability and performance.

$$z = \frac{x - \mu}{\sigma} \quad \text{(vi)}$$

Here x is the data point

Feature Engineering: Extraction or selection of features is an important part to improve the classification accuracy. Conventional techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are widely used for feature extraction and selection during the analysis of data. Principal component analysis retains most of the variance while reducing the dimensionality of the data in order to make the data easier for analysing and interpreting, as shown in eq (vii-xii) [31].

Compute the mean (vii)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Center the data (viii)

$$\tilde{X} = X - \bar{x}$$

Compute the covariance matrix (ix)

$$\Sigma = \frac{1}{n}\tilde{X}^T\tilde{X}$$

Eigenvalue decomposition (x)

$$\Sigma v_i = \lambda_i v_i$$

Here v_i are eigen vectors and λ_i are eigen values

Selection of Principal components (xi)

$$V_k = [v_1, v_2, \dots, v_k]$$

Transform the data (xii)

$$Y = \tilde{X} V_k$$

While as Linear Discriminant Analysis, selects the linear combinations of features that perfectly separates the classes which thereby improving classification performance [32]. In terms of advanced techniques, Convolutional Neural Networks (mainly convolution layer and pooling layer (eq (xiii-xv)) are skilled to extract the features in the hierarchical form from the image data which make them suitable for tasks like analysing medical images where it is important to identify and understand the patterns like edges and textures [33].

Convolution Operation

$$(I * K)(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x + i, y + j).K(i, j) \quad \text{(xiii)}$$

Here m, n are dimensions of filter K

Activation function

$$Z(x, y) = \sigma\big((I * K)(x, y) + b\big) \quad \text{(xiv)}$$

Where σ = activation function, b = bias

Pooling Operation

$$P(x, y) = max_{i,j}\, Z(s.x + i, s.y + j) \quad \text{(xv)}$$

Where s = stride length

Recurrent Neural Networks along with Long Short-Term Memory networks and Gated Recurrent Unit Network excel with sequential data and capture temporal dependencies in medical records such as progression of disease over time and predict patient outcomes. On the other hand, autoencoders are unsupervised networks that compress the data and then reconstruct it. Such technique is useful for dimensionality reduction of data and feature extracts features in complex medical and datasets in order to enhance the accuracy of diagnosis and treatment planning in healthcare applications [34].

Hybrid Model Training: Advanced Convolutional Neural Network and their hybridizing techniques play a vital role in the context of identifying pancreatic cancer. CNNs are basically designed for recognizing images and can improvise their efficiency in medical diagnostics on being hybridized with other neural networks. These techniques include integrating CNNs with pre trained models for effectively using large dataset as well as incorporate attention mechanisms which focus on important features present in the medical images. Moreover, hybridizing CNN with recurrent neural network efficiently improves the analysis of sequential data which later on enhances the classification or prediction rate [35].
In the domain of pancreatic cancer, advanced neural network systems are useful in assaying the medical image data as its detection in early stage is essential to improve the clinical outcomes of patient outcomes. Apart from this, these networks also assist to detect normal abnormalities by spontaneously learning hierarchical representations to indicate early stages of pancreatic cancer. Moreover, hybrid approaches like CNN-RNN combinations allow the examination of time-series data from medical records which aid in monitoring disease course and plan treatment accordingly [36].

Optimization: Hyperparameter optimization is essential in the development of Convolutional Neural Networks for predicting pancreatic cancer due to several key reasons. Firstly, CNNs consist of numerous hyperparameters such as learning rates, batch sizes, and network architectures (e.g., number of layers, filter sizes) that significantly impact model performance. Fine tuning the hyperparameters with the optimization techniques can enhance the ability of CNN in extracting the relevant features from medical images. This helps in the improvisation of accuracy to identify subtle signs of pancreatic cancer [37]. Secondly, the process of hyperparameter optimization, such as Grid Search, Hyperband, Genetic, or Bayesian Optimization,

ensures that these settings are systematically explored as they help to identify configurations that produce the best prediction performance of the model without modifying the parameters in order to save effort as well as time in the development of model. Furthermore, hyperparameter optimization is also useful in reducing the risk of modelling errors such as overfitting or underfitting or misclassification effectively and enables the model to generalize better to new as well as unseen medical images which thereby enhance their applicability and reliability in clinical settings [38].

Evaluation Metrics: Metrics such as accuracy, loss, precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC) provide an in-depth assessment of performance of CNN model to predict pancreatic cancer (eq (xvi-xxi)). Accuracy measures the overall correctness of predictions while as loss measures the deviation between the actual and predicted values. Precision quantifies the percentage of true effective predictions amongst all true predictions, recall assesses the share of actual positives correctly diagnosed, and the F1-rating balances precision and recall measure right into a single metric. AUC-ROC evaluates the classifier's capacity to differentiate among training, offering a summary of its discrimination ability throughout distinctive decision thresholds. Together, those metrics provide comprehensive information of how perfectly a CNN model performs in detecting pancreatic cancer, assisting in its optimization and clinical software [39].

$$Acc = \frac{True\ Positive + True\ Negative}{True\ Positive + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad \text{(xvi)}$$

$$\text{Loss} = \frac{(\text{Actual Value} - \text{Predicted Value})^2}{\text{Number of observations}} \quad \text{(xvii)}$$

$$\text{Prec} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad \text{(xviii)}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad \text{(xix)}$$

$$\text{F1 score} = 2\frac{\text{Precision} * \text{Recall}}{\text{Recall} + \text{Precision}} \quad \text{(xx)}$$

$$\text{AUC} - \text{ROC} \approx \sum_{i=1}^{N-1} \frac{1}{2}(\text{FPR}_{i+1} - \text{FPR}_i).(\text{TPR}_{i+1} + \text{TPR}_i \quad \text{(xxi)}$$

Here, FPR = false positive rate and TPR = true positive rate

Explainability (XAI): Techniques like LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), etc are valuable tools for interpreting model predictions, especially in complex models including CNNs for the prediction of pancreatic cancer. SHAP, as shown in eq (xxii), presents explanations through computing the contribution of each characteristic to the model's prediction across all possible subsets of capabilities. This approach is rooted in cooperative game principle, mainly Shapley values, which attribute the prediction final results to each feature considering its interaction with different functions. By quantifying the effect of each function on predictions, SHAP offers insights into how specific enter variables affect the outcome of models [40].

$$\emptyset_i = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{i\}} \frac{|S|!\,(p - |S| - 1)!}{p!} [f(S \cup \{i\}) - f(S)] \quad \text{(xxii)}$$

Here, S is subset of features, p = total features, f(S) = prediction for subset S

LIME, however, generates local factors by means of approximating the predictions of a black-box model together with a CNN, using interpretable models (e.g, linear models) trained on multiple times of the

authentic data. By specializing in nearby regions around precise predictions, LIME provides insights into why a model made a specific selection for an individual, helping in knowledge model behavior at a granular level. Both SHAP and LIME contribute to model interpretability, helping researchers and clinicians recognize the elements influencing CNN predictions for pancreatic cancer [40]. These techniques are crucial for ensuring transparency, trustworthiness, and medical attractiveness of AI-pushed diagnostic equipment in medical applications.

RQ3: What are the advantages of Collaborative studies and accessible data programs in advancing deep learning approaches for pancreatic cancer detection?

Collaborative studies and accessible data programs play crucial roles in advancing deep learning approaches for pancreatic cancer detection, as shown in Figure 6.

Figure 6: Roles of collaborative approaches and open data initiatives

These initiatives allow researchers to access different datasets including different imaging techniques such as MRI, CT scans, histopathological images and patient profile. Such diversity helps in training deep learning models that are generalizable and robust across various conditions and population. The accessibility of larger datasets by means of collaboration and open data initiatives provides more occurrences of pancreatic cancer cases as well as healthy controls. This enriches the stability of deep learning models by enabling better training and validation which thereby enhances the reliability as well as performance of the model. Moreover, collective resource allocation and expertise through collaboration quickens the speed of innovation in developing new deep learning algorithms and techniques for pancreatic cancer detection. This collaborative effort often results in progress that individual researchers or institutions might not gain independently [41]. Collaborative research additionally streamlines the foundation of benchmarks as well as requirements for assessing the performance of deep learning models in figuring out pancreatic cancer. This standardization is important to compare multiple processes and make sure constant assessment metrics throughout studies [42]. Open statistics projects or accessible records programs promote validation research throughout a couple of institutions and affected person populations supporting to evaluate the reliability and generalizability of deep learning networks for pancreatic cancer detection beneath awesome medical settings as well as conditions. Furthermore, collaborative studies generally involve multidisciplinary teams incorporating specialists from pathology, oncology, computer technology, radiology, as well as bioinformatics. This interdisciplinary method stimulates the innovative synthesis of thoughts and methodologies which result in progressive answers that fuse various viewpoints [43]. Open information tasks inspire transparency in research by making datasets and methodologies openly handy, improving the reproducibility of study's findings and taking into account evaluation of ethical considerations associated with records consent, privacy, and research conduct. Finally, collaborative studies initiatives encourage community engagement and educational outreach. They

provide opportunities for clinicians, researchers, patients, and the public to take part in and contribute to advancements in pancreatic most cancers detection thru sharing of statistics, discussions, and know-how propagation. By promoting collaborative and large research surroundings, those projects pave the way for massive improvements inside the detection and treatment of pancreatic cancer [44].

5. Conclusion

The review paper focused on the employment of AI learning models that has reshaped the way of pancreatic cancer detection by providing unique capabilities to analyze any type of medical data. In order to enhance accuracy as well as sensitivity of pancreatic cancer, these techniques outperformed to extract intricate patterns while being taken directly from raw data. AI models, in spite of their promising results also face limitations in terms of sample size issues, small datasets, interpretability challenges, and disparities in validation methods. To address these, the paper suggested the incorporation of optimized advanced Convolutional Neural Networks with XAI to achieve higher reliability in the identification of cancer. Apart from this, the review paper also underlined the value of integrating imaging data with laboratory or clinical data to analyse the risk and provide customized diagnostic perception of pancreatic cancer using AI techniques. In addition to this, it also emphasized the capability of combining traditional techniques with AI methods in order to boost the capabilities of diagnosing which further lead to optimized early detection, rigorous classification, and personalized therapy planning for overall quality care of patient. The future research direction involves the implementation of these AI techniques in clinical settings to improvise interpretation of models and remarkably escalate the accuracy graph of identifying pancreatic cancer.

No conflict of interest

References

1. Puckett Y, Garfield K. Pancreatic Cancer. StatPearls - NCBI Bookshelf 2022. https://www.ncbi.nlm.nih.gov/books/NBK518996/.
2. Facts About Pancreatic Cancer. American Cancer Society 2024. https://www.cancer.org/cancer/types/pancreatic-cancer/about/key-statistics
3. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. Ca 2024;74:12–49. https://doi.org/10.3322/caac.21820.
4. Tripathi S, Tabari A, Mansur A, Dabbara H, Bridge CP, Daye D. From Machine Learning to Patient Outcomes: A Comprehensive Review of AI in Pancreatic Cancer. Diagnostics. 2024 Jan 12;14(2):174. https://www.mdpi.com/2075-4418/14/2/174
5. Goggins M. The role of biomarkers in the early detection of pancreatic cancer. Familial Cancer. 2024 Apr 25:1-4. https://link.springer.com/article/10.1007/s10689-024-00381-4
6. Chen X, Hu X, Liu T. Development of liquid biopsy in detection and screening of pancreatic cancer. Frontiers in Oncology. 2024;14. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11180763/
7. Gupta A, Koul A, Kumar Y. Pancreatic cancer detection using machine and deep learning techniques. In2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM) 2022 Feb 23 (Vol. 2, pp. 151-155). IEEE. https://ieeexplore.ieee.org/abstract/document/9754010/
8. Tests and Procedures Used to Diagnose Cancer. CancerGov 2023. https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis.
9. Gupta S, Kumar Y. Cancer prognosis using artificial intelligence-based techniques. SN Computer Science. 2022 Jan;3(1):77. https://link.springer.com/article/10.1007/s42979-021-00964-3
10. Koul A, Kumar Y, Gupta A. A study on bladder cancer detection using AI-based learning techniques. In2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS) 2022 Oct 10 (pp. 600-604). IEEE. https://ieeexplore.ieee.org/abstract/document/9988462/
11. Hameed BS, Krishnan UM. Artificial intelligence-driven diagnosis of pancreatic cancer. Cancers. 2022 Oct 31;14(21):5382. https://www.mdpi.com/2072-6694/14/21/5382
12. Viriyasaranon T, Chun JW, Koh YH, Cho JH, Jung MK, Kim SH, Kim HJ, Lee WJ, Choi JH, Woo SM. Annotation-Efficient Deep Learning Model for Pancreatic Cancer Diagnosis and Classification Using CT Images: A Retrospective Diagnostic Study. Cancers. 2023 Jun 28;15(13):3392. https://doi.org/10.3390/cancers15133392

13. Acer I, Bulucu FO, Içer S, Latifoğlu F. Early diagnosis of pancreatic cancer by machine learning methods using urine biomarker combinations. Turkish Journal of Electrical Engineering and Computer Sciences. 2023;31(1):112-25. https://journals.tubitak.gov.tr/elektrik/vol31/iss1/8/
14. Muhammad W, Hart GR, Nartowt B, Farrell JJ, Johung K, Liang Y, Deng J. Pancreatic cancer prediction through an artificial neural network. Frontiers in Artificial Intelligence. 2019 May 3;2:2. https://doi.org/10.3389/frai.2019.00002
15. Jan Z, El Assadi F, Abd-Alrazaq A, Jithesh PV. Artificial intelligence for the prediction and early diagnosis of pancreatic cancer: Scoping review. Journal of Medical Internet Research. 2023 Mar 31;25:e44248. https://doi.org/10.2196/44248
16. Dinesh MG, Bacanin N, Askar SS, Abouhawwash M. Diagnostic ability of deep learning in detection of pancreatic tumour. Scientific Reports. 2023 Jun 15;13(1):9725. https://www.nature.com/articles/s41598-023-36886-8
17. Huang B, Huang H, Zhang S, Zhang D, Shi Q, Liu J, Guo J. Artificial intelligence in pancreatic cancer. Theranostics. 2022;12(16):6931. https://doi.org/10.7150/thno.77949
18. Nasief H, Zheng C, Schott D, Hall W, Tsai S, Erickson B, Allen Li X. A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer. NPJ precision oncology. 2019 Oct 4;3(1):25. https://www.nature.com/articles/s41698-019-0096-z
19. Luo Y, Chen X, Chen J, Song C, Shen J, Xiao H, Chen M, Li ZP, Huang B, Feng ST. Preoperative prediction of pancreatic neuroendocrine neoplasms grading based on enhanced computed tomography imaging: validation of deep learning with a convolutional neural network. Neuroendocrinology. 2020 Apr 29;110(5):338-50. https://doi.org/10.1159/000503291
20. Lee KS, Jang JY, Yu YD, Heo JS, Han HS, Yoon YS, Kang CM, Hwang HK, Kang S. Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: Retrospective cohort study. International Journal of Surgery. 2021 Sep 1;93:106050. https://www.sciencedirect.com/science/article/pii/S1743919121001849
21. Ramaekers M, Viviers CGA, Hellström T a. E, Ewals LJS, Tasios N, Jacobs I, et al. Improved Pancreatic Cancer Detection and Localization on CT Scans: A Computer-Aided Detection Model Utilizing Secondary Features. Cancers 2024;16:2403. https://doi.org/10.3390/cancers16132403.
22. Shi W, Wartmann T, Accuffi S, Al-Madhi S, Perrakis A, Kahlert C, et al. Integrating a microRNA signature as a liquid biopsy-based tool for the early diagnosis and prediction of potential therapeutic targets in pancreatic cancer. British Journal of Cancer 2023;130:125–34. https://doi.org/10.1038/s41416-023-02488-4.
23. Dodda K, Muneeswari G. IANFIS: a machine learning–based optimized technique for the classification and segmentation of pancreatic cancer. Research on Biomedical Engineering 2024. https://doi.org/10.1007/s42600-024-00352-9.
24. Ravi KC, Dixit RR, Singh S, Gopatoti A, Yadav AS. AI-powered pancreas navigator: Delving into the depths of early pancreatic cancer diagnosis using advanced deep learning techniques. In2023 9th International Conference on Smart Structures and Systems (ICSSS) 2023 Nov 23 (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/10407836.
25. Placido D, Yuan B, Hjaltelin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nature Medicine 2023;29:1113–22. https://doi.org/10.1038/s41591-023-02332-5.
26. Tonozuka R, Itoi T, Nagata N, Kojima H, Sofuni A, Tsuchiya T, et al. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. Journal of Hepato-Biliary-Pancreatic Sciences 2020;28:95–104. https://doi.org/10.1002/jhbp.825.
27. Mahapatra S. Why Deep Learning over Traditional Machine Learning? Medium 2019. https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063
28. Taye MM. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation. 2023 Mar 6;11(3):52. https://www.mdpi.com/2079-3197/11/3/52
29. Liao Y, Liu H, Spasić I. Deep learning approaches to automatic radiology report generation: A systematic review. Informatics in Medicine Unlocked. 2023 Jan 1;39:101273.
30. Helwan A, Ma'aitah MK, Hamdan H, Ozsahin DU, Tuncyurek O. Radiologists versus Deep Convolutional Neural Networks: A Comparative Study for Diagnosing COVID-19. Computational and Mathematical Methods in Medicine. 2021;2021(1):5527271. 10.1155/2021/5527271

31. What Is Principal Component Analysis (PCA)? | IBM n.d. https://www.ibm.com/topics/principal-component-analysis
32. Dash SK. A Brief Introduction to Linear Discriminant Analysis. Analytics Vidhya 2024. https://www.analyticsvidhya.com/blog/2021/08/a-brief-introduction-to-linear-discriminant-analysis
33. Kumar Y, Bansal S, Jariyal A, Koul A. Artificial intelligence-based learning techniques for accurate prediction and classification of colorectal cancer. CRC Press eBooks, 2024, p. 114–9. https://doi.org/10.1201/9781003471059-16.
34. Ravindran U, Gunavathi C. A survey on gene expression data analysis using deep learning methods for cancer diagnosis. Progress in Biophysics and Molecular Biology. 2023 Jan 1;177:1-3. https://www.sciencedirect.com/science/article/abs/pii/S0079610722000803
35. Koul A, Bawa RK, Kumar Y. An analysis of deep transfer learning-based approaches for prediction and prognosis of multiple respiratory diseases using pulmonary images. Archives of Computational Methods in Engineering. 2024 Mar;31(2):1023-49. https://link.springer.com/article/10.1007/s11831-023-10006-1
36. Udriștoiu AL, Cazacu IM, et al. Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. PLoS One. 2021 Jun 28;16(6):e0251701. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0251701
37. Kumar Y, Garg P, Moudgil MR, Singh R, Woźniak M, Shafi J, Ijaz MF. Enhancing parasitic organism detection in microscopy images through deep learning and fine-tuned optimizer. Scientific Reports. 2024 Mar 8;14(1):5753. https://www.nature.com/articles/s41598-024-56323-8
38. Ali YA, Awwad EM, Al-Razgan M, Maarouf A. Hyperparameter search for machine learning algorithms for optimizing the computational complexity. Processes. 2023 Jan 21;11(2):349. https://www.mdpi.com/2227-9717/11/2/349
39. Koul A, Bawa RK, Kumar Y. Enhancing the detection of airway disease by applying deep learning and explainable artificial intelligence. Multimedia Tools and Applications. 2024 Feb 21:1-33. https://link.springer.com/article/10.1007/s11042-024-18381-y
40. Srinidhi B, Bhargavi MS. An XAI Approach to Predictive Analytics of Pancreatic Cancer. In2023 International Conference on Information Technology (ICIT) 2023 Aug 9 (pp. 343-348). IEEE. https://ieeexplore.ieee.org/abstract/document/10225991
41. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, Folio LR, Summers RM, Rubin DL, Lungren MP. Preparing medical imaging data for machine learning. Radiology. 2020 Apr;295(1):4-15. https://pubs.rsna.org/doi/full/10.1148/radiol.2020192224
42. Dewey M, Bosserdt M, Dodd JD, Thun S, Kressel HY. Clinical imaging research: higher evidence, global collaboration, improved reporting, and data sharing are the grand challenges. Radiology. 2019 Jun;291(3):547-52. https://doi.org/10.1148/radiol.2019181796
43. Whitney HM, Baughan N, Myers KJ, Drukker K, Gichoya J, Bower B, Chen W, Gruszauskas N, Kalpathy-Cramer J, Koyejo S, Sá RC. Longitudinal assessment of demographic representativeness in the Medical Imaging and Data Resource Center open data commons. Journal of Medical Imaging. 2023 Nov 1;10(6):061105-. https://doi.org/10.1117/1.JMI.10.6.061105
44. Althobaiti MM, Almulihi A, Ashour AA, Mansour RF, Gupta D. [Retracted] Design of Optimal Deep Learning-Based Pancreatic Tumor and Nontumor Classification Model Using Computed Tomography Scans. Journal of Healthcare Engineering. 2022;2022(1):2872461. https://doi.org/10.1155/2022/2872461
45. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE transactions on medical imaging. 2019 Jan 23;38(8):1777-87. https://doi.org/10.1109/TMI.2019.2894349