Explainable AI for Pancreatic Cancer Prediction and Survival Prognosis: An Interpretable Deep Learning and Machine Learning Approach

Srinidhi B, M S Bhargavi*
Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India
E-mail: srinidhib18@gmail.com, ms.bhargavi@gmail.com
*Corresponding Author

Keywords: explainable artificial intelligence, shapley additive explanations (SHAP), local interpretable model-agnostic explanations (LIME), deep learning, pancreatic cancer

Received: September 2, 2023

Pancreatic cancer's devastating impact and low survival rates call for improved detection methods. While Artificial Intelligence has shown remarkable progress, its increasing complexity has led to "black box" models, hindering their acceptance in critical fields like healthcare. To address this, Explainable Artificial Intelligence (XAI) has gained traction, aiming to create transparent AI systems. In this study, we propose a comprehensive approach that combines the power of Deep Learning for pancreatic cancer detection using Computed Tomography (CT) images and Machine Learning (ML) for survival prognosis based on clinical data. By leveraging CT images with Deep learning models such as Convolutional Neural Networks, VGG-16 and DenseNet-201, effective diagnosis of Pancreatic Cancer is achieved and comprehensive insights into the tumor's spatial characteristics are obtained. DenseNet-201 outperformed the other models in terms of accuracy and interpretability with a predictive accuracy of 95%. The integration of ML techniques such as Stochastic Gradient Descent, Naïve Bayes and Extra Tree classifiers with clinical data predicts the chances of survival, providing vital information for treatment planning and personalized care. To validate the model's accuracy and interpretability, a comprehensive XAI validation is conducted using state-of-the-art techniques like Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations. These methods provide localized explanations for predictions, allowing clinicians to understand risk and survival chances. This study holds immense potential to aid healthcare professionals in diagnosis, prognosis, and personalized treatment strategies, contributing to enhanced patient outcomes in the fight against pancreatic cancer.

Povzetek: Analizirana sta napovedovanje in prognoza preživetja pri raku trebušne slinavke z uporabo globokega učenja in razložljive umetne inteligence (XAI) za interpretacijo napovedi, izboljšanje zaupanja v AI.

1 Introduction

Pancreatic cancer is a devastating malignancy that poses a significant global health challenge due to its aggressive nature and poor prognosis [1]. Despite advances in medical research and cancer treatments, the five-year survival rate for pancreatic cancer patients remains dishearteningly low, with many cases being diagnosed at advanced stages when effective interventions become limited. Cancer detection and chance of survival are critical factors that can significantly impact patient outcomes by enabling timely treatments and personalized therapeutic strategies. Consequently, there is an urgent need to explore novel approaches that offer deeper insights into pancreatic cancer development and patient survival rates.

Pancreatic cancer's clinical landscape is characterized by its insidious nature in the development stages, often presenting vague or non-specific symptoms. This lack of distinct clinical indicators makes early diagnosis challenging, resulting in delayed treatment initiation and reduced chances of successful interventions. Additionally, the intricate biology of pancreatic cancer necessitates a comprehensive understanding of the underlying molecular and genetic factors driving disease progression and influencing patient outcomes [2]. As such, the development of accurate predictive models that can identify individuals at high risk of developing pancreatic cancer and estimate patient survival probabilities becomes imperative in the fight against this lethal disease.

While current medical technologies have significantly contributed to cancer detection and diagnosis, they still face limitations. Imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and endoscopic ultrasound provide valuable visualizations of pancreatic tumors and staging information. However, these imaging techniques may miss small lesions or produce ambiguous results in anatomically complex areas, leading to diagnostic

challenges. Furthermore, the use of biomarker analysis, including CA 19-9 and several other factors for monitoring treatment response and assessing disease progression may lack specificity and sensitivity, resulting in false-positive or false-negative outcomes. These limitations highlight the need for more accurate and comprehensive predictive tools for improved cancer detection and patient survival estimation.

Advancements in AI technology enabled the rapid and precise identification of cancer through diverse medical imaging methods. The Deep Learning (DL) algorithms have garnered significant attention in medical imaging analysis. DL algorithms, particularly Convolutional Neural Networks (CNNs), can automatically learn hierarchical representations from raw image data, such as CT scans. This unique capability enables DL models to detect complex patterns and features in medical images, offering great potential for aiding the detection and diagnosis of pancreatic cancer [3]. Additionally, deep learning can be integrated with clinical data, including patient demographics and medical history, to develop comprehensive predictive models that deliver more accurate and personalized survival estimates

Machine learning (ML) models can be trained on diverse clinical data to predict patient survival rates by considering factors such as age, tumor grade, CA 19-9 levels, sex, and medical history. By leveraging these ML-based approaches, clinicians can obtain individualized survival chances, which are essential for tailoring treatment plans and optimizing patient care. However, DL and conventional ML models often operate as black boxes, lacking transparency in their decision-making process. This opacity hinders the widespread adoption of DL and ML in critical clinical decision-making settings, where understanding the factors influencing predictions is paramount.

To address the interpretability challenge associated with DL and ML models, XAI techniques have emerged as a solution [4]. In this study, we propose a comprehensive approach that combines the power of Deep Learning for pancreatic cancer detection using CT images and Machine Learning for survival prognosis based on clinical data. To enhance the interpretability and trustworthiness of our model, we incorporate XAI techniques, specifically Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP).

LIME plays a crucial role in providing interpretable insights into the DL-based cancer detection model's predictions. By employing LIME, we can identify the specific regions within the CT images that significantly contribute to the model's decision-making process. These regions act as crucial markers for the presence of tumors and other abnormalities, enabling radiologists and clinicians to validate and understand the model's findings. Mapping CT image regions relevant to the model's predictions not only boosts confidence in AI-driven detection but also assists medical professionals in recognizing subtle signs of pancreatic cancer, potentially enabling earlier diagnosis and timely intervention.

Additionally, SHAP plays a pivotal role in explaining machine learning-based survival rate predictions using clinical data. SHAP provides a quantitative measure of the influence of each clinical feature on the model's survival rate estimations. By identifying the most affected features, SHAP empowers healthcare professionals to understand the factors driving the model's predictions and their relative importance in determining patient survival probabilities. This level of interpretability allows clinicians to prioritize critical clinical factors and consider them in treatment planning and patient management. SHAP's ability to reveal the most influential clinical factors further enhances the personalized nature of our model's survival predictions, ensuring tailored and optimized patient care.

By integrating LIME and SHAP into our interpretable analytical model, we provide a comprehensive framework that not only accurately detects pancreatic cancer using CT images but also offers transparent and interpretable survival rate predictions based on clinical data. The combination of these XAI techniques not only bolsters the trustworthiness of the model's predictions but also empowers medical professionals to make informed decisions based on AI-driven insights. Ultimately, our model's enhanced interpretability contributes to improved patient outcomes by facilitating earlier detection, more precise prognostication, and personalized treatment strategies for individuals battling pancreatic cancer.

The remainder of the paper is organized as follows. Section I provides the Introduction to the research. Section II presents the literature review of existing systems. Section III provides Dataset description. Section IV presents the proposed methodology and implementation details. Section V presents the results and discussion of the proposed system. Section VI concludes the work with future scope.

2 Related work

Pancreatic cancer remains a formidable and deadly malignancy, necessitating innovative approaches to improve early detection and prognostication for enhanced patient outcomes. The recent convergence of DL for cancer detection using CT images, ML for survival rate prediction based on clinical data, and the integration of XAI techniques have shown great potential in tackling the challenges presented by this devastating disease.

The incorporation of XAI methods, such as LIME and SHAP, enhances the interpretability and trustworthiness of AI-driven models, empowering medical professionals to comprehend the factors driving predictions and make well-informed clinical decisions. In this literature survey, we delve into research focused on the development and application of interpretable analytical models for pancreatic cancer detection and survival prognosis. By examining the advances and insights from these studies, we aim to provide a comprehensive understanding of the state-of-the-art techniques that hold the potential to revolutionize

pancreatic cancer management and contribute to improved patient care.

The recent convergence of DL techniques for cancer detection using CT images has shown significant promise in addressing the challenges posed by this devastating disease. Among the various DL approaches, CNNs and their extensions, such as recurrent CNNs, have emerged as particularly effective in the detection of pancreatic cancer. These methods have demonstrated impressive accuracy in identifying pancreatic tumors and differentiating them from normal tissues, as evidenced by multiple studies [3,5,6,7,8,9,10,11]. CNNs are well-suited for processing the complex and high-dimensional data found in medical imaging, enabling them to extract intricate patterns and features that are indicative of cancerous tissues. The effectiveness of these models is largely attributed to their ability to learn from large datasets, which enhances their capacity to recognize subtle distinctions between malignant and benign tissues. Recurrent CNNs, which integrate temporal and spatial information, have further advanced the field by improving the detection accuracy and robustness of these models. This approach allows for the consideration of sequential dependencies in image slices, which is crucial for the accurate identification of tumors in CT scans. The application of DL models offers significant advantages over traditional diagnostic methods, including faster processing of large volumes of image data and higher accuracy in tumor detection. As a result, these advancements hold the potential to facilitate earlier diagnosis, thereby improving treatment outcomes.

To address the interpretability challenge associated with DL and ML models, XAI techniques have emerged as a solution. The incorporation of XAI methods, such as LIME and SHAP enhances the interpretability and trustworthiness of AI-driven models, empowering medical professionals to comprehend the factors driving predictions and make well-informed clinical decisions. To better understand the model's decision-making process, a limited number of studies have explored the use of LIME and SHAP for predictions. LIME, in particular, offers valuable insights by mapping regions on CT images and highlighting specific areas that influence the model's predictions and an attempt for such analysis has been made in [12]. In some research, LIME has been utilized for feature importance analysis in predicting Pancreatic cancer [13,14,15].

Survival prediction in pancreatic cancer aids in improved treatment planning, provides realistic prognostic information for patients and their families, and supports the clinical trial design and assessment of new therapies. Research on predicting patient survival rates for pancreatic cancer using machine learning models based on clinical data has revealed promising results, demonstrating the potential to significantly improve patient outcomes. Various ML algorithms, such as random forests, support vector machines, logistic regression, nearest neighbor and gradient boosting, have been explored and assessed for their performance in survival prediction [16,17]. These models utilize extensive clinical datasets to identify patterns and correlations that might not be apparent to human clinicians. The effectiveness of these models in survival prediction is attributed to their capacity to process and analyze complex data, offering more accurate prognostic insights than traditional methods.

Despite these advancements, the lack of interpretability in these models poses a significant challenge, hindering their seamless integration into clinical practice and limiting clinician trust. To address this limitation XAI methods like SHAP have been incorporated [12,18,19]. The incorporation of SHAP was emphasized to elucidate the influential clinical factors driving survival predictions, aiding clinicians in making data-driven decisions for personalized treatment planning.

The literature review reveals that there have been limited research efforts focused on developing interpretable analytical models for pancreatic cancer detection and survival prognosis. The application of interpretability in this field is still in its early stages, with only a limited number of studies available. Table 1 provides a summary of related works that utilize explainable AI models for pancreatic cancer prediction and survival prognosis. The integration of XAI techniques in these studies emphasizes the significance of transparency and trustworthiness in AI-driven medical applications. As these innovative methodologies continue to evolve, they hold the potential to revolutionize pancreatic cancer management and improve patient outcomes.

Another notable limitation observed across the surveyed papers is the absence of utilization of multiple models of DL and ML for interpretation and validation. The studies predominantly focused on applying specific XAI techniques to individual models, such as using LIME or SHAP for specific DL or ML models. Consequently, there was a lack of exploration into the advantages of employing a diverse set of DL and ML models for these purposes. This oversight hinders comprehensive comparison and validation of XAI techniques across different model architectures. There is a requirement to consider adopting multiple, diverse DL and ML models in their XAI framework to gain a more holistic understanding of the predictive features and foster the standardization of XAI practices in pancreatic cancer detection and survival prediction.

In conclusion, the literature survey has shed light on the progress made in developing interpretable analytical models for pancreatic cancer detection and survival prediction. The combination of DL, ML, and XAI techniques holds significant potential for enhancing early detection, personalized treatment planning, and ultimately improving patient care.

Table 1: Summary of related works with explainability of models in pancreatic cancer prediction and survival prognosis.

Authors	Focus Area	Models	Dataset	Performance Metrics	XAI Technique	Key Findings
Goel et al. (2021) [15]	Pancreatic Cancer Detection	Logistic Regression, Adaboost, Neural Network, ensemble model	Gene Expression Omnibus-miRNA biomarkers	Sensitivity – 0.85,0.90,0.88, 0.95. Specificity – 0.98,0.94,0.96, 0.98.	LIME SHAP	Ensemble models with feature selection and improved diagnosis accuracy for pancreatic cancer detection
Srinidhi B et. al. (2023) [12]	Pancreatic Cancer Detection	CNN SVM	CT Images, Clinical Data from - The Cancer Imaging Archive	None	LIME SHAP	Combined XAI with CNN and SVM for model interpretability on prediction
Bobes-Bascarán et al. (2024) [13]	Pancreatic Cancer Detection	Decision Trees, Random Forest, XGBoost	Clinical Data from -The Cancer Genome Atlas Program	Accuracy - 0.66,0.54,0.66 Precision - 0.72,0.76,0.76 Recall - 0.66,0.54,0.66	LIME SHAP	Feature importance analysis for Pancreatic Cancer
Dimitris et al. (2021) [19]	Survival Prognosis	Optimal Classification Tree, XGBoost	Clinical data of 2,784 Patients	1-year AUC for OCT - 0.63 3-year AUC - 0.67	SHAP	SHAP was utilized to identify the best predictor
Keyl et al. (2022) [18]	Survival Prognosis	Random Survival Forest	Albumin, CT image, Radiomics data, Molecular data of 203 Cohort	C-index - 0.71	SHAP	Feature importance analysis for survival prognosis from multi-modal data

3 Dataset description

In this research, we utilized the Cancer Imaging Archive (TCIA) dataset [20], a comprehensive collection of medical images and clinical data, to develop an interpretable analytical model for pancreatic cancer detection and survival rate prediction. The dataset comprises 100 cancer-positive CT images and 100 cancer-negative CT images, offering a diverse set of cases for robust model training and evaluation. For cancer detection, we employed an 80-20% split, with 80% of the images used for training the model and 20% for testing its performance. Furthermore, for survival prognosis, we utilized seven critical clinical features, including age, gender, KI-67 Index, PFS Months, Creatine, CA 19/9 U/ml and tumor grade. Around 50 patient records were considered for this work to create a well-rounded prognostic model. Similar to the cancer detection model, the survival prediction model was trained on 80% of the clinical data and evaluated on the remaining 20% to ensure reliable and generalizable results.

4 Methodology

The proposed system in this research aims to develop an interpretable analytical model for pancreatic cancer detection and survival prognosis, leveraging a combination of CT images and clinical data. The preprocessing pipeline for the CT images involves initial RGB conversion, normalization, and conversion to an array format, ensuring standardized input for subsequent analysis. These preprocessed images are then fed to Deep Learning models such as CNN [21], VGG-16 [22] and

DenseNet-201 [23] to extract meaningful features and perform accurate cancer predictions. To enhance the interpretability of the deep learning model, the LIME algorithm is applied. LIME generates region marking for the CT images, highlighting specific areas influencing the model's predictions. This approach empowers radiologists and medical professionals to validate the model's decision process, instilling confidence in the AI-driven diagnostic tool and enabling more informed decision-making for early cancer detection.

In parallel, the clinical data is preprocessed through label encoding and the removal of irrelevant columns to ensure data quality and uniformity. This preprocessed clinical data is then utilized to train and evaluate different machine learning models, such as Stochastic Gradient Descent (SGD) [24], Extra Tree [25], and Naïve Bayes [26], for survival rate prediction. To ensure interpretability in the machine learning models' survival predictions, the SHAP algorithm is employed. SHAP provides insights into the most influential clinical features affecting the model's prognostic outcomes, empowering medical practitioners to make data-driven decisions and design personalized treatment plans.

The suggested system's architecture is depicted in Figure 1. At the initial level of the architecture, data processing occurs, providing input for the deep learning and machine learning models. The outcomes are then validated using explainable artificial intelligence techniques. For cancer prediction, the data preparation includes image conversion, normalization, and image-to-array. The pre-processed CT images will be fed to the DL models, and the predictions made by these models (CNN, VGG-16, DenseNet-201) will be validated using LIME. Similarly, the Machine learning models (Extra Tree, SGD, Naïve Bayes) are applied to the pre-processed clinical dataset. To validate the model's results, we have employed SHAP for the survival prognosis.

4.1 Data pre-processing

Before utilizing the dataset for model development, pre-processing steps were undertaken to ensure data compatibility and optimize model performance. For CT images, a series of pre-processing steps were applied. First, to facilitate feature extraction, the images were converted from the native DICOM format to RGB (Red-Green-Blue) representation. This conversion allowed the incorporation of color information, enabling the model to capture subtle patterns in the images effectively. Subsequently, to normalize the pixel values and bring them within a standardized range, a normalization [27] process was performed. This step ensures that the model is not overly influenced by variations in pixel intensity across different images, promoting consistent and reliable predictions. Finally, the processed CT images were transformed into arrays, facilitating their integration into the model.

Regarding the clinical data, appropriate pre-processing steps were undertaken to optimize its usability in the model. Categorical variables in the clinical data, such as gender or tumor grade, were subjected to label encoding [28] to convert them into numerical representations.

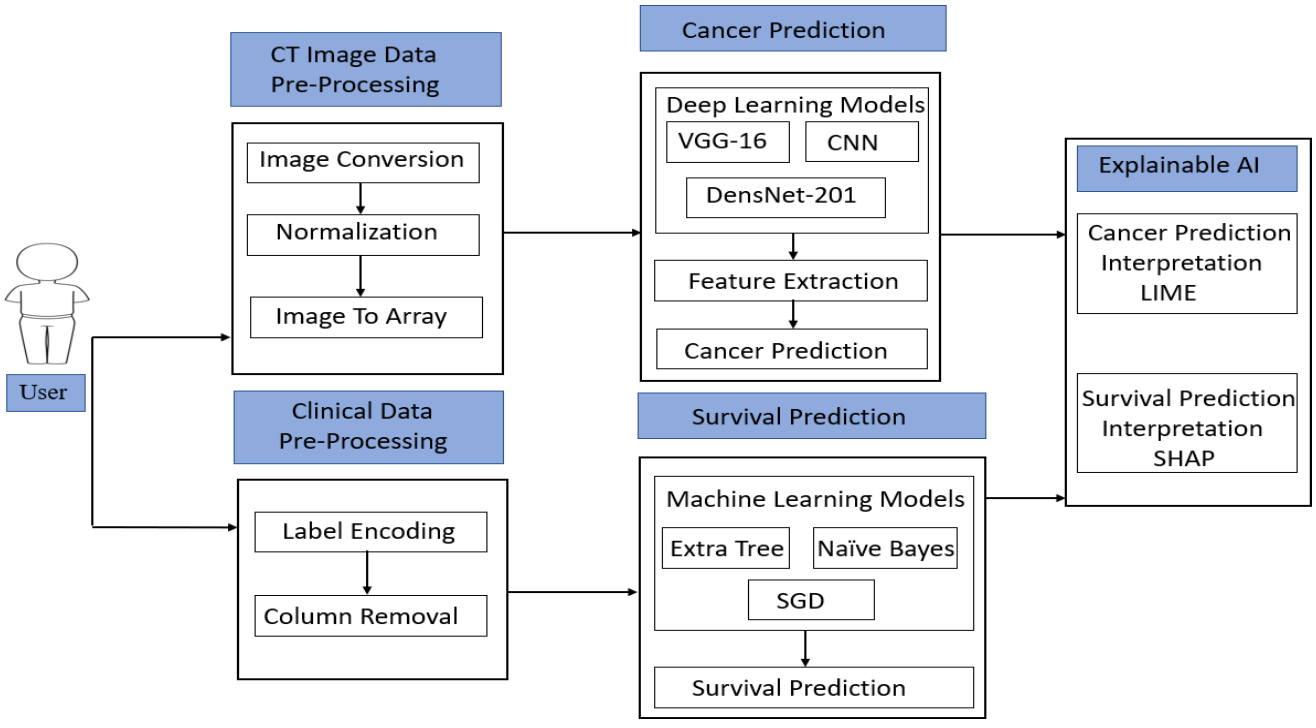

Figure 1: System architecture

This process enabled the model to work seamlessly with categorical data during the training and prediction phases. Additionally, to enhance model efficiency and eliminate irrelevant information, unwanted columns containing redundant or non-contributing features were removed from the clinical dataset. By streamlining the clinical data, the model can focus on relevant clinical factors that significantly influence the prognosis and improve the model's interpretability.

The pre-processing steps undertaken for both CT images and clinical data were essential to prepare the dataset for the subsequent model development. By ensuring data compatibility, standardization, and relevance, the interpretable analytical model can effectively leverage the combined information from CT images and clinical features to facilitate accurate pancreatic cancer detection and survival rate prediction, contributing to improved patient care and clinical decision-making.

4.2 Deep learning models and architecture

In this study, we have implemented a comprehensive framework for pancreatic cancer detection and survival prediction using state-of-the-art machine learning and deep learning models. For cancer prediction, we employed Deep learning models such as CNN, VGG-16 and DenseNet-201, which are well-suited for cancer prediction tasks involving medical imaging data like CT scans. Pancreatic cancer diagnosis relies heavily on accurately detecting subtle abnormalities in CT images. CNNs are adept at automatically learning complex features from images, making them effective in capturing intricate patterns indicative of cancerous regions in pancreatic CT scans.

A standard CNN represents a fundamental deep learning architecture, known for its simplicity and effectiveness in various image classification tasks. It serves as a baseline to compare the interpretability and performance of more complex models. VGG-16 and DenseNet-201 are deep architectures with many layers, allowing them to extract and combine rich hierarchical features, which are beneficial for the challenging task of pancreatic cancer detection. VGG-16 architecture is deeper and more sophisticated than a basic CNN. VGG-16 is a well-established model in the field of image recognition, known for its good performance and moderate complexity. DenseNet-201 is an advanced and highly deep architecture that includes dense connections between layers, enhancing feature propagation and their reuse.

By employing these three architectures, we can assess the effectiveness and consistency of interpretability methods across models of varying complexity. This enables us to understand whether these interpretability methods can provide reliable explanations for both simple and complex models, which is critical for their application in medical AI.

4.2.1 CNN model

In this study, we implemented Convolutional Neural Network model for pancreatic cancer detection using CT images as input and achieving binary yes/no predictions. The CNN architecture consists of two convolutional layers with a 3x3 filter size and 64 filters in each layer followed by a max-pooling layer with a pooling size of 2x2 to downsample the feature maps. To prevent overfitting, we applied dropout regularization with a dropout rate of 0.25 after each convolutional layer. The CNN architecture is shown in Figure 2 [21]. The learning rate was set to 0.001, and we used the Adam optimizer [29] to efficiently update the model's parameters during training.

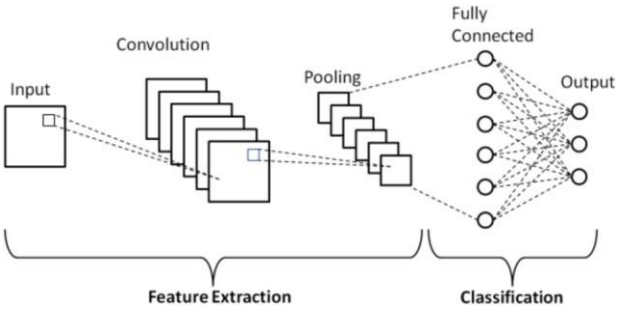

Figure 2: CNN architecture

Due to the limited size of the dataset, we decided to conduct a smaller number of epochs for training the CNN models. Specifically, we set the number of epochs to 20. The rationale behind this decision was that with a minimal dataset, the model tends to reach a saturation point in learning after a few epochs. Continuing training beyond this point led to overfitting, where the model performs well on the training data but fails to generalize to unseen data as depicted in Figure 3.

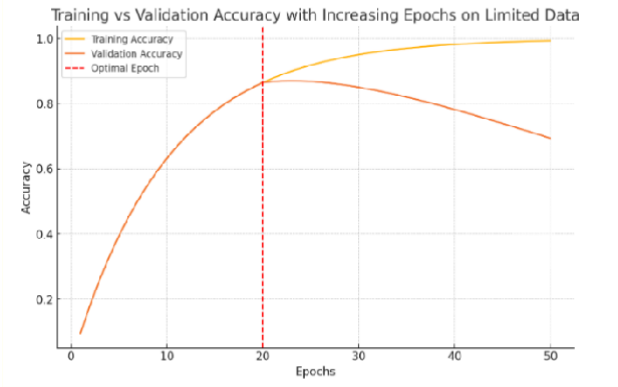

Figure 3: Epochs v/s Accuracy Graph

By fine-tuning the hyperparameters and training the CNN models with a reduced number of epochs, we aimed to strike a balance between achieving high accuracy on the training data and ensuring generalization performance on the testing data. This approach proved effective in managing the limited

dataset size and allowed us to develop reliable binary prediction models for pancreatic cancer detection using CT images as input.

4.2.2 VGG-16 model

The VGG-16 model architecture depicted in Figure 4 consists of 13 convolutional layers, each with a filter size of 3x3. The number of filters in each consecutive block varies, starting from 64 in the initial layers and progressively increasing to 128, 256, 512, and finally 512 in the deeper layers. Max-pooling is applied with a size of 2x2 to reduce spatial dimensions and extract relevant features effectively. To prevent overfitting, dropout regularization is incorporated with a rate of 0.5, randomly deactivating neurons during training to improve generalization. We used a learning rate of 0.001, a batch size of 32, and trained the model for 20 epochs. The decision to use 20 epochs was based on the limited size of the dataset and the observation that the error rate reached a stable value after these iterations. We used the Adam optimizer [29] to efficiently update the model's parameters during training.

The hyperparameters were carefully chosen to strike a balance between model performance and computational efficiency given the dataset constraints. Overall, the VGG-16 model implementation proved to be effective in generating binary yes/no predictions for pancreatic cancer detection, providing valuable insights into the potential of deep learning in early cancer diagnosis.

4.2.3 DenseNet-201

The DenseNet-201 architecture illustrated in Figure 5, is a deep convolutional neural network architecture for image recognition tasks. It employs a dense connectivity pattern where each layer receives direct input from all preceding layers, promoting feature reuse and enhancing gradient flow. The hyperparameters used for DenseNet-201 were carefully selected to achieve optimal performance. The growth rate was set to 32, ensuring the network's ability to capture relevant features effectively. We employed 6 blocks, each consisting of 48 layers, to create a deep and expressive architecture capable of capturing intricate patterns in the CT images.

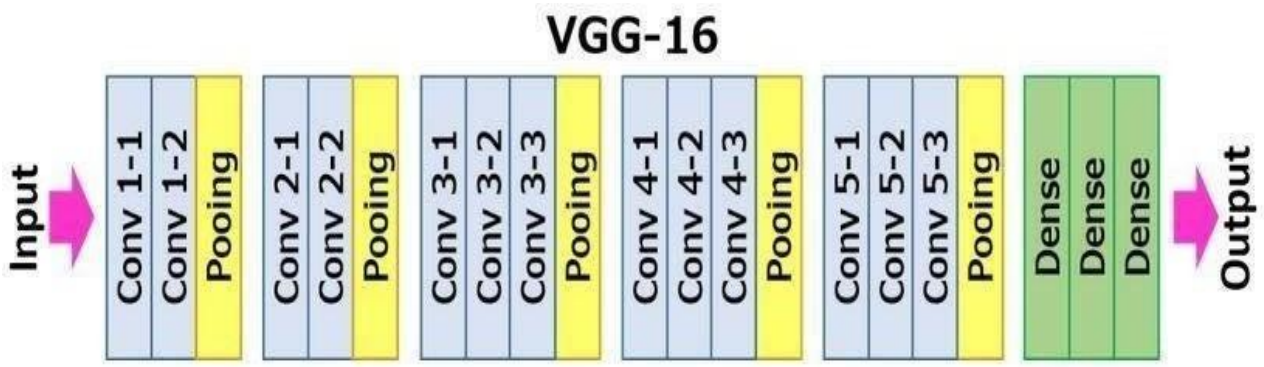

Figure 4: VGG architecture

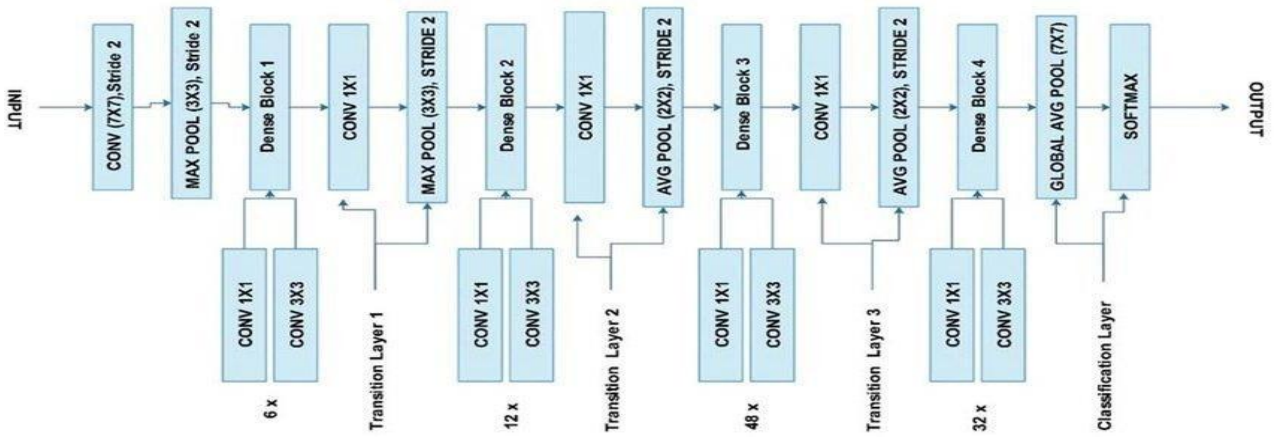

Figure 5: DenseNet-201 architecture

To prevent overfitting and improve generalization, we applied a compression factor of 0.5, which reduced the number of feature maps in transition layers. Additionally, dropout with a rate of 0.5 was introduced during training to regularize the model. We initialized the learning rate at 0.001 to facilitate a stable and effective optimization process. The choice of batch size was set to 32, balancing computational efficiency and model convergence. Considering the limited size of the dataset, we conducted training for 20 epochs, as observed that the error rate reached a near-constant level after this point. The model was trained over 20 epochs and for updating the model's parameters during the training process, Adam optimizer [29] was utilized.

In our study, hyperparameter tuning was critical for optimizing model performance. We employed grid search to systematically evaluate combinations of key hyperparameters, including learning rate, dropout rate, and batch size for all three models. Specifically, we tested learning rates of [0.0001, 0.001, 0.01], dropout rates of [0.25, 0.5], and batch sizes of [16, 32, 64]. The Adam optimizer was used across all models to ensure robust performance metrics. This approach helped us identify the optimal settings that minimized validation loss and maximized the accuracy of the models.

4.3 Machine learning models and architecture

Survival prediction in pancreatic cancer refers to the estimation of a patient's likelihood of surviving for a certain period after diagnosis, typically measured in months or years. This prediction is based on various factors, including the stage of cancer at diagnosis, patient demographics, tumor characteristics, treatment options, and other clinical data. For survival prognosis using clinical data, the choice of algorithms depends on the nature of the features and the sizeof the dataset. State-of-the-art Machine learning algorithms such as Stochastic Gradient Descent, Naïve Bayes and Extra Tree classifiers are utilized. The choice of these ML techniques is due to their diversity in model behavior.

4.3.1 Stochastic gradient descent classifier

Stochastic Gradient Descent is a widely used optimization algorithm suitable for large datasets with multiple features, as it can handle complex models effectively. This linear model is chosen for its simplicity and efficiency, particularly with large datasets. Its behavior is easy to interpret, which aligns well with the use of interpretability methods.

The goal is to estimate the chance of survival as either high or low for patients with pancreatic cancer. The SGD Regressor is a powerful machine learning algorithm that optimizes the model parameters iteratively to minimize the loss function, making it well-suited for regression tasks. To configure the SGD Regressor, we tuned several hyper-parameters to achieve the best performance. This included setting the learning rate, which controls the step size for parameter updates during training. A learning rate of 0.01 was selected to strike a balance between rapid convergence and avoiding overshooting the optimal solution. Additionally, we introduced regularization to prevent overfitting and enhance model generalization. A regularization strength of 0.1 was chosen to control the amount of regularization applied to the model. The number of iterations was set to 100 to allow the model to update its parameters gradually and converge to an optimal solution. Early stopping criteria were introduced to prevent overfitting and reduce training time. We set the early stopping threshold to 10 iterations without improvement in the loss function.

To control the type of regularization used, we applied theL2 penalty, which adds a regularization term to the loss function proportional to the square of the model weights. This helped prevent model complexity and improved generalization.

4.3.2 Naïve bayes classifier

The Naive Bayes classifier is a probabilistic algorithm that operates on the assumption of independence among the features. This means that it calculates the likelihood of each feature given the class label (chance of survival) and then combines these probabilities to determine the overall probability of survival. It can handle feature independence assumptions and is computationally efficient, making it suitable for pancreatic cancer survival prediction tasks.

Unlike other models that may involve adjusting learning rates or tuning the number of layers, the Naive Bayes classifier is inherently straightforward and computationally efficient. It is particularly useful when dealing with high-dimensional data, which can be the case in clinical datasets. As a probabilistic model, Naïve Bayes offers a different perspective compared to linear models. It assumes feature independence, making it useful for exploring how individual features contribute to predictions.

4.3.3 Extra tree classifier

Extra Tree, a type of decision tree, is advantageous when dealing with small to medium-sized datasets and categorical features. It is well-suited for pancreatic cancer survival prediction, as clinical data often includes categorical information, such as tumor grade and disease stage. This ensemble method provides a non-linear model that can capture complex interactions between features. It's particularly useful in scenarios where the relationships between features are not straightforward. Applying interpretability methods to this model allows for the exploration of how feature interactions influence predictions. The model utilized 100 estimators for ensemble learning, employed "auto" as the maximum number of features considered per split, used "Gini" as the criterion for splitting internal nodes, and allowed unlimited tree depth. It also required a minimum of 2 samples for node splitting and 1 sample for leaf nodes. With these well-optimized configurations, the Extra Tree classifier effectively predicted survival chances using

clinical data, aiding personalized treatment planning and providing reliable prognostic insights for pancreatic cancer patients.

4.4 Explainable artificial intelligence

Explainable Artificial Intelligence (XAI) [4] refers to a set of techniques and methods that aim to make complex AI models more transparent and interpretable to humans. The need for XAI arises because many advanced AI models, such as deep learning neural networks, are often considered "black boxes," [30] meaning their decision-making process is difficult for humans to understand. XAI techniques provide insights into how the model arrives at its predictions, enabling users, such as medical professionals, to trust and interpret the model's results.

For Pancreatic Cancer prediction and survival prognosis, LIME is specifically chosen for region mapping because it can help highlight the important image regions influencing the CNN model's cancer prediction. In the context of CT images for pancreatic cancer, LIME can identify the specific regions in the scan that are crucial in the CNN's decision-making process, allowing radiologists to verify the model's focus on potential cancerous areas. This transparency aids in understanding the model's strengths and limitations, contributing to improved trust and confidence in the AI-driven diagnostic tool.

On the other hand, SHAP is employed to determine the most affected features in the machine learning models used for survival prediction. SHAP provides a global explanation of the model's feature importance, which is valuable in understanding how specific clinical factors impact the model's prognostic outcomes. For pancreatic cancer, SHAP can reveal which clinical features, such as age, tumor grade, or stage, significantly influence the model's prediction of patient survival rates. This information empowers clinicians to focus on key factors in personalized treatment planning and patient care, making SHAP an essential XAI technique for survival prognosis in pancreatic cancer management.

4.4.1 Local interpretable model-agnostic explanation

The deep learning model's output, which represents the cancer likelihood of each CT image, served as the input for the LIME algorithm. For each image in the dataset, LIME generated perturbations by introducing small changes to the input image while keeping the rest of the image unchanged. After creating these perturbations, we passed them through the deep learning model to obtain their corresponding cancer likelihood predictions. LIME then weighed these perturbations based on their proximity to the original image and the similarity of the predictions. By constructing a local linear model that approximates the behavior of the complex deep learning model in the vicinity of the input image, LIME extracted the coefficients to determine the impact of different regions of the CT image on the final cancer likelihood prediction. Through this process, LIME generated region mapping by highlighting the specific areas in the input CT image that had the most significant impact on the deep learning model's prediction. These regions of interest provided valuable insights into the image features and patterns influencing the model's decision-making process for cancer detection. The LIME-generated region mapping enhanced the interpretability of the deep learning model, allowing medical professionals to validate and gain confidence in its performance for pancreatic cancer detection. This transparent and interpretable explanation facilitated a more informed and reliable diagnostic process, ultimately contributing to improved patient care in the realm of pancreatic cancer management.

4.4.2 Shapley additive explanation

For each machine learning model prediction, we use the SHAP algorithm to explain the model's output. SHAP provides us with insights into the contribution of each clinical feature to survival prediction. By quantifying the impact of each feature, SHAP helps us identify the most influential clinical parameter affecting the patient's chance of survival. In the interpretation stage, the SHAP values for each feature are plotted or presented in descending order of importance. This visualization allows medical professionals to quickly grasp the most significant clinical factor influencing survival prediction. Furthermore, the SHAP values can be used to generate summary plots or individual feature importance profiles, enabling a comprehensive understanding of how different clinical parameters contribute to the overall survival prognosis. The integration of SHAP in our machine learning model ensures that we not only obtain accurate survival predictions but also gain valuable insights into the key clinical factors driving these predictions [31]. This interpretable approach empowers healthcare practitioners to make informed decisions and tailor treatment strategies for individual patients, ultimately enhancing the overall patient care and management of pancreatic cancer.

5 Experimental analysis and results

5.1 Experimental setup

The system is designed to provide accurate cancer prediction using deep learning models CNN, VGG-16 and DenseNet-201, as well as survival prediction using machine learning models SGD, Extra Tree, and Naïve Bayes. The input data consists of CT images for cancer prediction and clinical data for survival prediction. To ensure model interpretability, the system incorporates LIME and SHAP algorithms. The front-end interface is built using HTML and CSS, allowing users to interact with the system seamlessly. The integration of the models is achieved through a Flask application, enabling easy deployment and accessibility. The coding and development of the models are carried out using Jupyter Notebook for efficient prototyping and experimentation. All the experimental cases are developed in Python in a congested environment using Anaconda tools. The

competing classification approach and various feature extraction techniques are also used, and the system is configured with an Intel Core i5-6200U processor runningat 2.30 GHz and 8GB of RAM.

5.2 Interpretation of CNN model

The CNN model effectively identified the CT image as cancer-positive with an accuracy of 0.92, indicating a potential malignancy. To interpret and validate the CNN model's prediction, LIME generated region markings on the CT image, highlighting the specific areas that influenced the model's decision, thereby offering transparency and interpretability as shown in Figure 6. The marked regions indicated the regions of interest that the model relied upon to classify the image as cancer positive. The region mapping may indicate the presence of abnormal or suspicious structures within the pancreas or surrounding tissues, such as tumors, lesions, or other cancerous features. These marked regions could correspond to distinct patterns, shapes, or textures that the CNN model learned to associate with cancerous areas in the CT images. Similarly, the CNN model effectively identified the CT image as cancer-negative, indicating a potential malignancy. To interpret and validate the CNN model's prediction, LIME generated region markings on the CT image, highlighting the specific areas that influenced the model's decision, thereby offering transparency and interpretability.

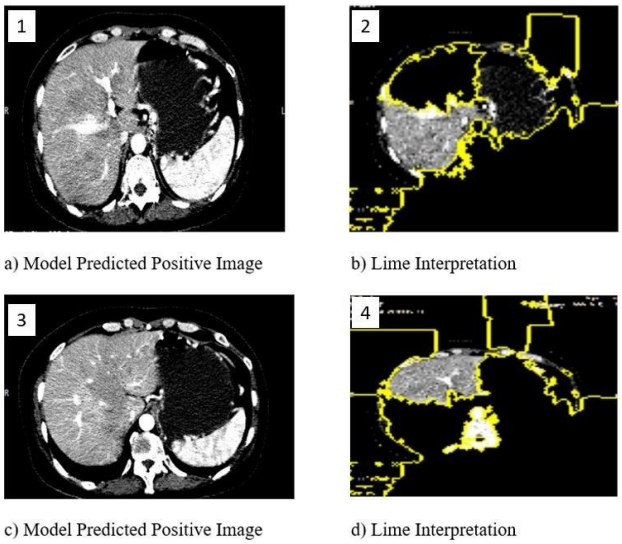

Figure 6: CNN interpretation for positive image

5.3 Interpretation of DenseNet-201 model

The interpretation provided by the DenseNet-201 model in region mapping using the LIME algorithm is potentially superior to that of a traditional CNN model. This is due to DenseNet's unique architecture and skip connections which enable more efficient feature propagation and deeper layer utilization. DenseNet-201 is a densely connected neural network that incorporates skip connections, allowing direct connections between all layers within the network. This dense connectivity facilitates the propagation of gradients and information throughout the network, leading to better feature reuse and representation. In the context of region mapping, this means that the DenseNet-201 model can capture a more comprehensive and intricate understanding of the CT images' features, including subtle patterns and contextually relevant information. As a result, when LIME performs region mapping using the DenseNet-201 model, it can identify and highlight even more specific and relevant regions within the CT images thatinfluenced the model's prediction. The dense connectivity and feature reuse in DenseNet-201 allow for more precise localization of the regions of interest, potentially offering a more detailed and accurate interpretation of the areas that are indicative of cancerous regions. Similarly, the DenseNet-201 model effectively identified the CT image as cancer-negative, indicating a potential malignancy. To interpret and validate the DenseNet-201 model's prediction, LIME generated region markings on the CT image, highlighting the specific areas that influenced the model's decision, thereby offering transparency and interpretability.

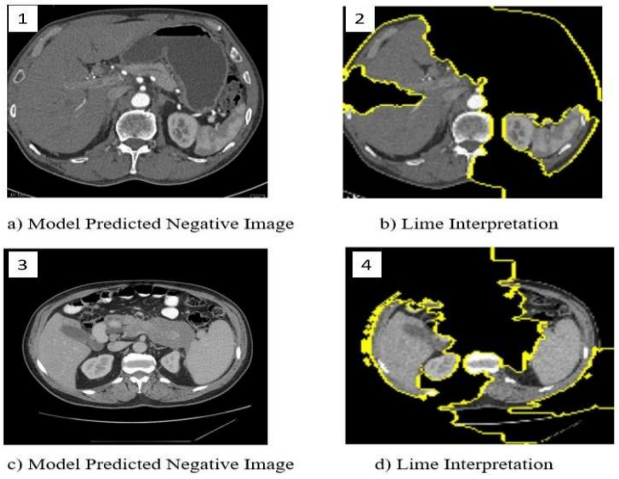

Figure 7: CNN interpretation for negative image

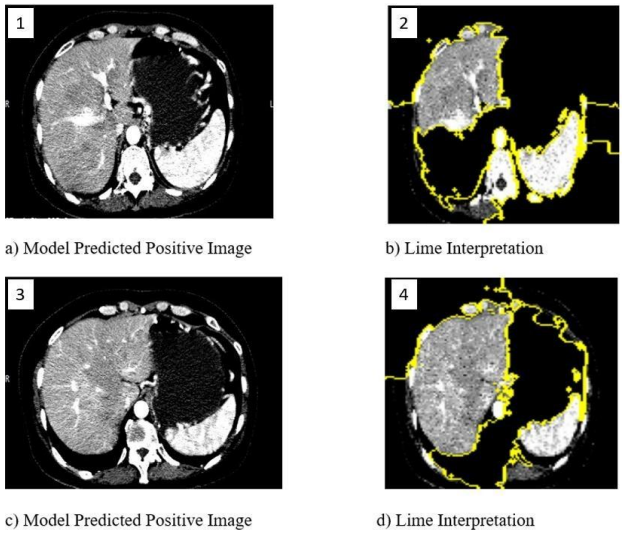

Figure 8: DenseNet-201 interpretation for positive image

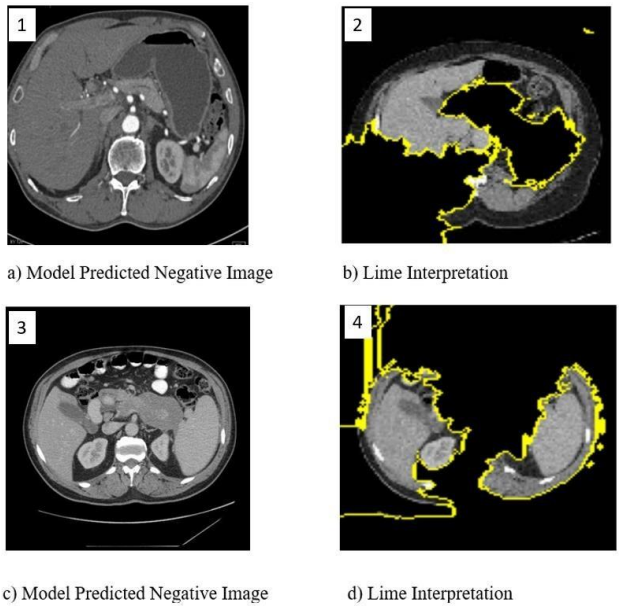

Figure 9: DenseNet-201 interpretation for negative image

5.4 Interpretation of VGG-16 model

The VGG-16 model demonstrated commendable predictive capabilities, accurately identifying the CT image as cancer-positive, which highlights its proficiency in distinguishing cancerous regions from non-cancerous ones. However, when comparing VGG-16's interpretability in region mapping through the LIME algorithm with the results obtained from DenseNet-201, we observed that VGG-16 exhibited some limitations. While VGG-16 effectively provided relevant region markings, it showed relatively poorer results compared to DenseNet-201 in precisely localizing intricate patterns and subtle features associated with pancreatic cancer.

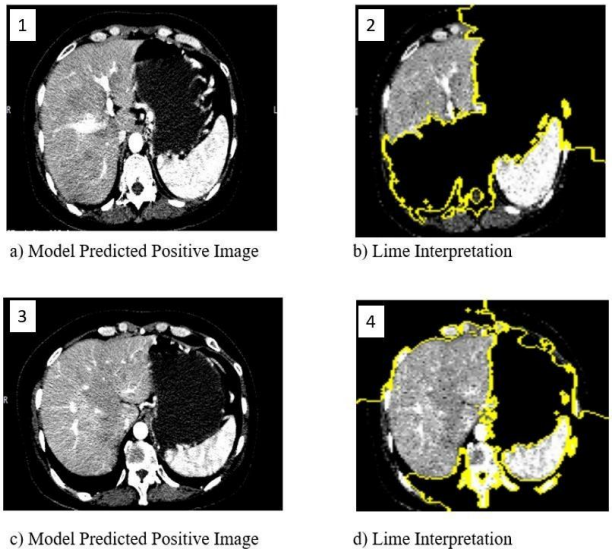

Figure 10: VGG-16 interpretation for positive image

The difference in performance can be attributed to the architectural dissimilarity between VGG-16 and DenseNet-201. DenseNet-201's densely connected structure allows for better feature reuse and a more holistic understanding of the CT images, enabling a more detailed and accurate region mapping through LIME. In contrast, VGG-16's deeper architecture with a higher number of parameters may have led to limited feature reuse and potentially diminished sensitivity to specific cancerous regions. Similarly, the VGG-16 model effectively identified the CT image as cancer-negative, indicating a potential malignancy. To interpret and validate the VGG-16 model's prediction, LIME generated region markings on the CT image, highlighting the specific areas that influenced the model's decision, thereby offering transparency and interpretability.

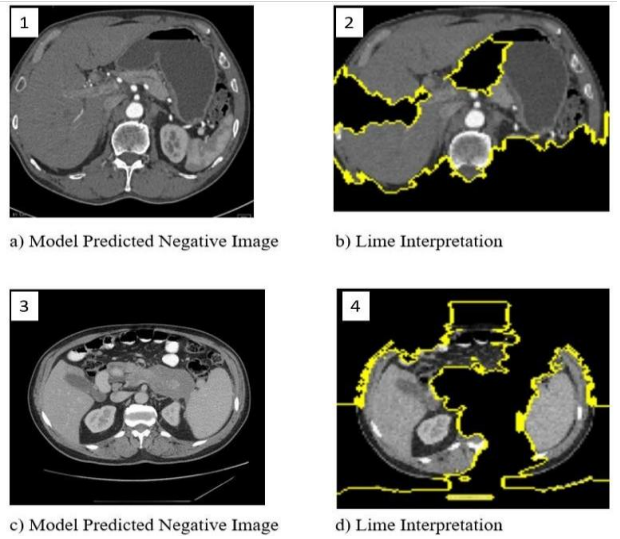

Figure 11: VGG-16 interpretation for negative image

Overall, in the context of pancreatic cancer detection using CT images, the DenseNet-201 model outperformed both the CNN and the VGG-16 models. DenseNet-201 demonstrated superior predictive capabilities, providing accurate and reliable cancer-positive predictions. Additionally, DenseNet-201 showed better interpretability through the LIME algorithm, delivering more precise and detailed region mappings compared to VGG-16 and CNN. The dense connectivity and feature reuse of DenseNet-201 enabled it to capture intricate patterns and contextually relevant information in CT images, leading to a more comprehensive and accurate localization of cancerous regions.

LIME provided the best region mapping results when combined with the DenseNet-201 model. The superior region mapping with LIME and DenseNet-201 can be attributed to the unique architecture of DenseNet-201, which facilitates better feature reuse and representation. This, in turn, allowed LIME to identify the most relevant and influential regions in the CT images, providing a comprehensive view of the features contributing to cancer prediction. The precise region mapping obtained from LIME with DenseNet-201 enhanced interpretability, instilling greater trust in the model's predictions, and enabling medical professionals

to validate and comprehend the decision-making process with confidence.

The model's significance lies in its potential to revolutionize pancreatic cancer management in society. By combining the accurate and interpretable predictions from DenseNet-201 and LIME, our model provides medical practitioners with a reliable AI-driven diagnostic tool. The model's transparency and interpretability through LIME facilitate better understanding and validation of predictions, fostering more informed clinical decision-making. Additionally, the efficient integration of deep learning and XAI techniques optimizes diagnostic workflows, reducing medical costs, and improving access to timely and accurate cancer diagnosis for a broader population. Ultimately, this will improve patient outcomes, supporting medical professionals in their mission to combat pancreatic cancer, and making a positive impact on society's healthcare landscape.

Evaluation methods: In the performance assessment of our experiment, we use four crucial evaluation metrics: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). True Positive represents the number of cases correctly predicted as positive or cancerous instances. False Positive indicates the number of cases wrongly predicted as positive when they are actually negative or non-cancerous. True Negative reflects the number of cases correctly predicted as negative or non-cancerous instances. False Negative, on the other hand, represents the number of cases incorrectly predicted as negative when they are actually positive or cancerous. These four measurements are fundamental in understanding the model's predictive accuracy and its capability to correctly classify cancer-positive and cancer-negative instances in pancreatic cancer detection

Accuracy = (TP+TN) / (TP+TN+FP+FN)
Precision = TP / (TP+FP)
Recall = TP / (TP+FN)
F1 Score = 2× (Precision *Recall) / (Precision +Recall)

Table 2: Performance analysis

DL Model	Accuracy	Precision	Recall	F1-Score
CNN	0.92	0.92	0.91	0.93
VGG-16	0.93	0.92	0.95	0.93
DenseNet-201	0.95	0.93	0.97	0.96

The comparison of accuracy among the three algorithms is provided in Table 2. DenseNet-201 emerges as the most accurate model for pancreatic cancer detection using CT images, achieving an impressive accuracy of 0.95. DenseNet-201's dense connectivity and skip connections allow it to effectively capture and reuse features, leading to a more comprehensive understanding of the intricate patterns indicative of pancreatic cancer. In close competition, VGG-16 also demonstrates strong performance with an accuracy of 0.93, showcasing its capability to distinguish cancerous and non-cancerous regions in CT images. While the CNN model performs well with an accuracy of 0.92, it falls slightly behind DenseNet-201 and VGG-16 in predictive power. These results highlight the superiority of deep learning architectures, particularly DenseNet-201, in accurately identifying pancreatic cancer using CT images, reinforcing their potential as valuable tools in the early diagnosis and management of this challenging disease.

Further, to determine the statistical significance of the differences in the predictive accuracies of the three models, one-way ANOVA statistical significance test is conducted. The ANOVA (Analysis of Variance) test checks whether there are statistically significant differences between the means of the different groups (in this case, CNN, DenseNet-201, and VGG-16). ANOVA returns a F-statistic of 6.2419 and p-value of 0.0199. F-statistic of 6.2419 indicates a notable variance between the models' accuracy (CNN, VGG-16, DenseNet-201) compared to the variance within each group. The p-value of 0.0199, being less than 0.05, confirms that the differences in accuracy between the models are statistically significant overall. Further analysis through post-hoc tests, such as Tukey HSD is performed, to pinpoint which specific pairs of models differ. Tukey HSD results are shown in Table 3.

According to Tukey's HSD test, DenseNet-201 has a significantly higher accuracy compared to CNN ($p = 0.0166$), while there is no significant difference between CNN and VGG-16 ($p = 0.4065$) or between DenseNet-201 and VGG-16 ($p = 0.1333$). This suggests that DenseNet-201 is the best-performing model among the three.

The graph in Figure 12 illustrates the Receiver Operating Characteristic (ROC) curves for three different models: CNN, VGG-16, and DenseNet-201. Each curve effectively plots the true positive rate against the false positive rate, showcasing the models' performance across various classification thresholds. The area under the curve (AUC) is indicated for each model, with CNN achieving an AUC of 0.52, VGG-16 at 0.51, and DenseNet-201 at 0.57.

These values demonstrate that all three models exhibit competitive discriminatory power, with DenseNet-201 showing a slight edge over the others. The results indicate that the models can provide valuable insights into the classification task at hand. The ROC curves serve as an effective visual tool for assessing model performance, highlighting the potential for further refinement and optimization of these models to enhance their predictive capabilities in future work.

Table 3: Results of Tukey HSD test

Group1	Group2	Mean Difference	P-adj	Lower	Upper	Reject
CNN	DenseNet-201	0.0325	0.0166	0.0066	0.0584	True
CNN	VGG-16	0.0125	0.4065	-0.0134	0.0384	False
DenseNet-201	VGG-16	-0.02	0.1333	-0.0459	0.0059	False

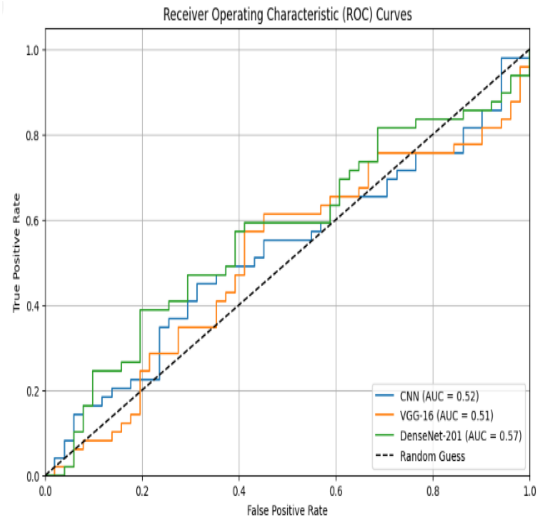

Figure 12: Receiver operating characteristic curve

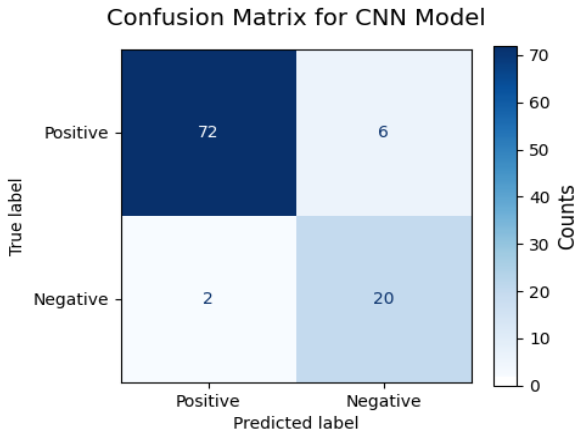

Figure 13: Confusion matrix for CNN.

The confusion matrices for the CNN, VGG-16 and DenseNet-201 are shown in Figures 13, 14 and 15 respectively providing insights into their classification performance. Each matrix shows the counts of true positives, true negatives, false positives, and false negatives for the respective models. These values help assess the accuracy and errors made during prediction, with true positives and negatives indicating correct classifications, while false positives and negatives highlight misclassifications.

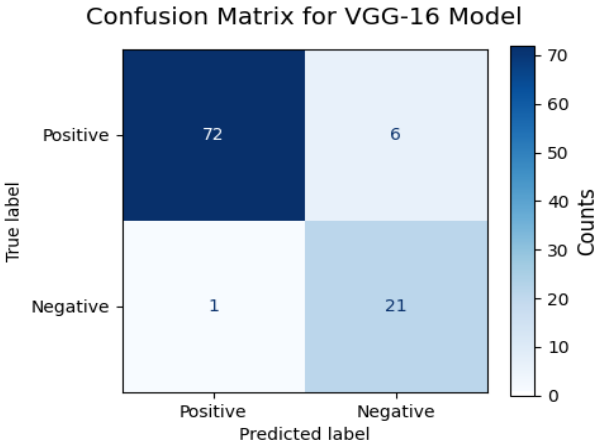

Figure 14: Confusion matrix for VGG-16.

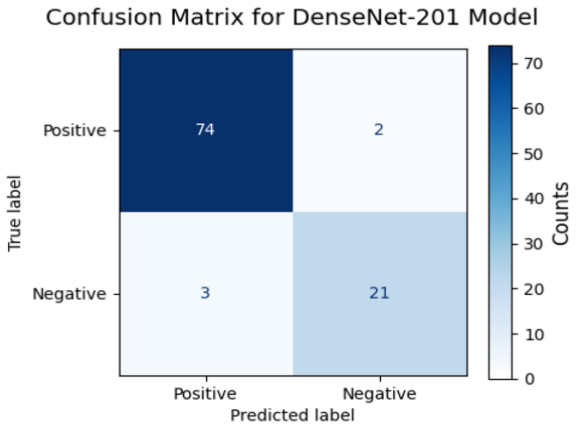

Figure 15: Confusion matrix for DenseNet-201

5.5 Analysis of survival prognosis

The selection of features for survival prediction in pancreatic cancer is a crucial step in building an accurate and meaningful model. The chosen features KI-67 Index, PFS Months, Creatine, Age, Tumor Grade, CA 19/9 U/ml, and Gender, have been carefully selected based on their clinical relevance and potential impact on patient survival outcomes [32].

KI-67 Index: KI-67 is a protein that is closely associated with cell proliferation and tumor growth. Its measurement provides valuable information about the rate of tumor cell division, which is a critical factor in predicting tumor aggressiveness and patient survival.

Table 4: Selected features

KI-67 Index
PFS Months
Creatine
Age
Tumor Grade
CA 19/9 U/ml
Gender

PFS months (Progression-Free Survival Months): PFS is a critical clinical endpoint in cancer prognosis, representing the time from the start of treatment to disease progression or relapse. It is a strong indicator of treatment efficacy and overall survival.

Creatine: Creatine levels in the blood can be indicative of kidney function, which is essential for monitoring potential complications and overall health status during cancer treatment.

Age: Age is a significant prognostic factor in cancer outcome. Younger patients may have a better overall health status and tolerate treatments more effectively, while older patients may have additional comorbidities that influence their survival.

Tumor grade: Tumor grade is a measure of tumor cell differentiation and aggressiveness. Higher tumor grades typically indicate more aggressive tumors that may have a poorer prognosis.

CA 19/9 U/ml: CA 19-9 is a tumor marker associated with pancreatic cancer. Elevated levels may indicate advanceddisease and a higher risk of poor outcomes.

Gender: Gender can also play a role in cancer prognosis, as some types of cancers may behave differently in males and females.

The selected features encompass a wide range of clinical and biological factors that are known to impact pancreatic cancer survival outcomes. By incorporating these diverse and relevant features, the survival prediction model can capture the complex interactions between clinical characteristics and tumor biology, leading to more accurate and personalized prognostic assessments. This feature selection process ensures that the developed model is both clinically meaningful and robust, enhancing its usefulness in guiding treatment decisions and improving patient outcomes in pancreatic cancer management.

While the other two algorithms, namely SGD and Naive Bayes, may also provide insights into the most affected features for survival prediction in pancreatic cancer, their accuracy in determining these features is not as precise as that of the Extra Tree classifier. The less accurate identification of the most influential clinical factors in the survival prognosis by SGD and Naive Bayes might be attributed to their respective algorithmic limitations. SGD is an optimization-based method that may not fully capture intricate feature interactions, while Naive Bayes assumes independence between features, potentially overlooking complex relationships. As a result, their feature importance rankings might lack the precision exhibited by the Extra Tree model. Due to this disparity in performance, we have chosen to focus on the Extra Tree classifier as it provides a more reliable and accurate understanding of the critical clinical factors influencing patient survival in pancreatic cancer.

The three models (SGD, Naive Bayes, and Extra Tree) may not necessarily give the same "most affected feature" in survival prediction using the SHAP algorithm. The reason lies in the inherent differences in the algorithms' working principles and the way they determine feature importance.

SGD is an optimization algorithm that iteratively updates the model's parameters to minimize the prediction error. It may assign different weights to different features during this process, leading to variations in feature importance.

Naive Bayes is a probabilistic algorithm based on the Bayes theorem. It assumes independence between features given the class label. Due to this assumption, Naive Bayes may not capture complex interactions between features, and its feature importance ranking might differ from other models.

Extra Tree is an ensemble learning method that builds multiple decision trees and combines their predictions. It selects random subsets of features and nodes during tree construction, introducing randomness that can lead to different feature importance rankings.

As a result, the three models may prioritize features differently in terms of their impact on survival predictions. The differences could be more pronounced when the dataset is not large enough to provide a comprehensive view of feature interactions. Comparing the models' performance metrics, such as feature importance scores derived from the SHAP algorithm, on anindependent test dataset would help identify the model that best captures the crucial clinical factors affecting patient survival in the context of survival prognosis. Additionally, conducting feature importance analysis and clinical validation with domain experts could shed more light on the models' interpretability and reliability in survival prognosis.

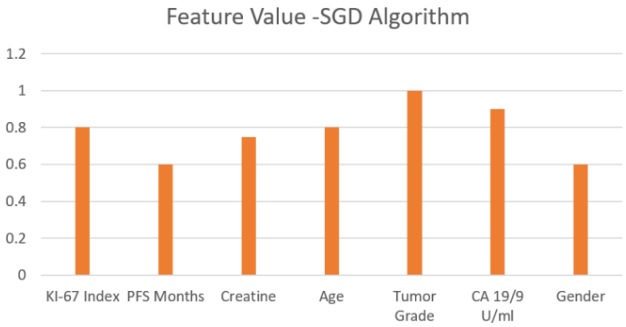

Figure 16: Feature value graph for SGD algorithm

Figure 16 displays feature importance values for the SGD algorithm in predicting survival outcomes for pancreatic cancer patients. Tumor grade holds the highest importance with a value of 1.0, followed by CA 19/9 U/ml (0.9) and KI-67 Index (0.8). Age, Creatine, and

PFS Months show importance values of 0.8, 0.7, and 0.6, respectively.

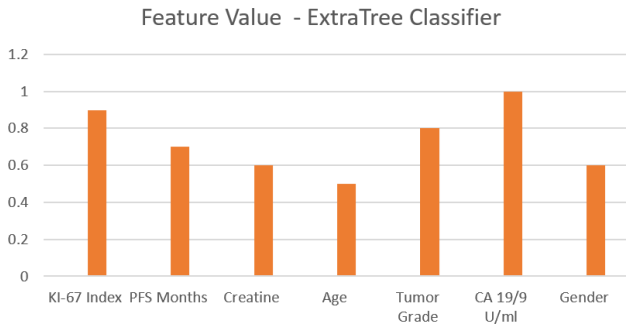

Figure 17: Feature value graph for extratree algorithm

The ExtraTree algorithm's feature importance graph as shown in Figure 17 illustrates the relative significance of each feature in predicting survival outcomes for pancreatic cancer patients. CA 19/9 U/ml emerges as the most important feature with a value of 1.0, signifying its critical role in determining patient survival. KI-67 Index closely follows with an importance value of 0.85, underscoring its considerable impact on survival predictions. Tumor Grade ranks third with a value of 0.8, indicating its significant contribution to survival prognosis. Additionally, PFS Months and Creatine, demonstrate importance values of 0.7 and 0.6, respectively, highlighting their relevance in predicting patient outcomes. Age exhibits the lowest importance with a value of 0.45, suggesting its relatively minor influence on survival predictions.

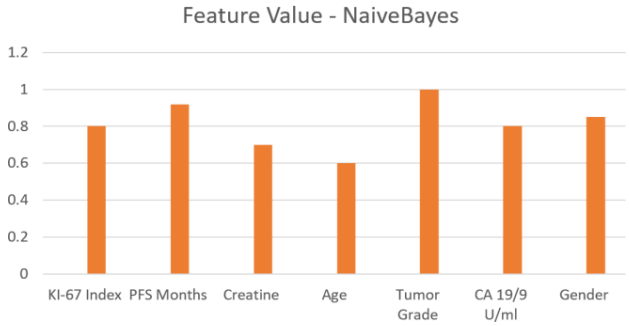

Figure 18: Feature value graph for naïve bayes algorithm

The feature importance graph as shown in Figure 18 for the Naive Bayes algorithm reveals the relative significance of each feature in predicting survival outcomes for pancreatic cancer patients. Tumor grade holds the highest importance with a value of 1.0, followed by PFS Months (0.85), CA 19/9 U/ml (0.8) and KI-67 Index (0.8). Age, Creatine, and PFS Months show lower importance scores.

In conclusion, based on the comparison of feature importance and interpretability using the SHAP algorithm, the Extra Tree classifier stands out as the top-performing model for survival prediction in pancreatic cancer. Its ability to accurately identify the most affected features highlights its potential as a valuable tool for medical professionals in personalized treatment planning and decision-making for patients with pancreatic cancer.

6 Conclusion and future work

In this comprehensive study, we investigated a diverse range of algorithms for pancreatic cancer prediction and survival prognosis. Our findings demonstrate that the VGG-16 and DenseNet-201 outperformed CNN algorithms in cancer prediction, achieving high accuracy in distinguishing cancer-positive and cancer-negative CT images. The adoption of LIME for model validation provided interpretable insights by marking the regions influencing the models' predictions, fostering trust and facilitating the integration of AI-driven diagnostics into clinical workflows.

For survival prediction, the machine learning models SGD, Extra Tree, and Naïve Bayes exhibited promising performance, with the SHAP algorithm delivering crucial insights into the most impactful clinical factors affecting patient outcomes. SHAP's ability to explain model predictions enabled medical practitioners to identify key prognostic indicators, empowering them to make informed decisions and improve patient care.

The utilization of Explainable Artificial Intelligence techniques, such as LIME and SHAP, ensured that both cancer prediction and survival prognosis were transparent and interpretable. This interpretability not only boosts the confidence of medical professionals in the model's predictions but also provides a deeper understanding of the intricate decision-making processes underlying each model. The study has significant implications for society by combining powerful AI-driven algorithms with interpretable XAI techniques, medical practitioners can confidently make data-driven decisions, leading to improved patient outcomes and survival rates.

Future work should focus on expanding the dataset to include more diverse cases and increasing the sample size to improve model generalization. Additionally, integrating other relevant data sources, such as genomic data or histopathological features, could lead to a more comprehensive and accurate cancer prediction system. Further research could explore the combination of multiple imaging modalities and clinical data to gain a deeper understanding of the disease. Moreover, incorporating other XAI techniques or developing hybrid interpretability methods might provide even greater insights into model predictions, boosting confidence and facilitating widespread clinical adoption. Finally, conducting rigorous validation studies on real-world patient data and collaborating with medical practitioners for clinical validation will be crucial for the successful translation of this framework into clinical practice.

Disclosure Statement

Statements of ethical approval

Not applicable

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profitsectors.

Competing Interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials / Code:
Data and code will be made available on request.

References

[1] K. Honda, Y. Hayashida, T. Umaki, T. Okusaka, T. Kosuge, S. Kikuchi, and F. Moriyasu. Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res., vol. 65, no. 22, pp. 10613-10622. 2005.

https://doi.org/10.1158/0008-5472.can-05-1851

[2] P. Maisonneuve and A. Lowenfels. Epidemiology of Pancreatic Cancer: An Update. Digestive Diseases (Basel, Switzerland), vol. 28, pp. 645-656. 2010.

https://doi.org/10.1159/000320068

[3] Si, Ke, Ying Xue, Xiazhen Yu, Xinpei Zhu, Qinghai Li, Wei Gong, Tingbo Liang, and Shumin Duan. "Fully end-to-end deep-learning-based diagnosis of pancreatic tumors." Theranostics 11, no. 4: 1982, 2021.

https://doi.org/10.7150/thno.52508

[4] F. K. Došilović, M. Brčić and N. Hlupić. Explainable artificial intelligence: A survey. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia. pp. 0210-0215, 2018.

http://dx.doi.org/10.23919/MIPRO.2018.8400040

[5] Wei Xuan, Guang qiang You. Detection and diagnosis of pancreatic tumor using deep learning-based hierarchical convolutional neural network on the internet of medical things platform. Future Generation Computer Systems, Volume 111, pp. 132-142, 2020. ISSN 0167-739X. https://doi.org/10.1016/j.future.2020.04.037.

[6] Z. Zhang, S. Li, Z. Wang, and Y. Lu. A Novel and Efficient Tumor Detection Framework for Pancreatic Cancer via CT Images. 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, pp. 1160-1164, 2020.

https://doi.org/10.1109/EMBC44109.2020.9176172

[7] Wei Zhao, Liyue Shen, Bin Han, Yong Yang, Kai Cheng, Diego A.S. Toesca, Albert C. Koong, Daniel T. Chang, Lei Xing. Marker less Pancreatic Tumor Target Localization Enabled by Deep Learning. International Journal of Radiation Oncology*Biology*Physics, Volume 105, Issue 2, pp. 432-439, 2019, ISSN 0360-3016.

https://doi.org/10.1016/j.ijrobp.2019.05.071.

[8] B. S. Hameed and U. M. Krishnan. Artificial Intelligence- Driven Diagnosis of Pancreatic Cancer. Cancers, vol. 14, no. 21, pp. 5382, 2022.

https://doi.org/10.3390/cancers14215382

[9] Li, M., Nie, X, Reheman, Y., Huang, P., Zhang, S., Yuan, Y., & Han, W. Computer-aided diagnosis and staging of pancreatic cancer based on CT images. IEEE Access, 8, pp. 141705-141718, 2020. https://doi.org/10.1109/access.2020.3012967

[10] D. Agarwal, O. Covarrubias-Zambrano, S. H. Bossmann and B. Natarajan. Early Detection of Pancreatic Cancers Using Liquid Biopsies and Hierarchical Decision Structure. IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, pp. 1-8, 2022.

https://doi.org/10.1109/jtehm.2022.3186836

[11] Sekaran, K., Chandana, P., Krishna, N. M., & Kadry, S.. Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. Multimedia Tools and Applications, 79 (15), pp. 10233-10247, 2020.

https://doi.org/10.1007/s11042-019-7419-5

[12] Srinidhi, B., and M. S. Bhargavi. An XAI Approach to Predictive Analytics of Pancreatic Cancer. International Conference on Information Technology (ICIT), pp. 343-348, 2023.

https://doi.org/10.1109/icit58056.2023.10225991

[13] Bobes-Bascarán, José, Eduardo Mosqueira-Rey, Ángel Fernández-Leal, Elena Hernández-Pereira, David Alonso-Ríos, Vicente Moret-Bonillo, Israel Figueirido-Arnoso, and Yolan da Vidal-Ínsua. Evaluating Explanatory Capabilities of Machine Learning Models in Medical Diagnostics: A Human-in-the-Loop Approach. arXiv preprint arXiv:2403.19820, 2024

[14] Bobes-Bascarán, José, Ángel Fernández-Leal, Eduardo Mosqueira-Rey, David Alonso Ríos, Elena Hernández-Pereira, and Vicente Moret-Bonillo. Understanding Machine Learning Explainability Models in the context of Pancreatic Cancer Treatment. VI Congreso Xove TIC: impulsando el talento científico. A Coruña, pp. 175-182. Universidade da Coruña, Servizo de Publicacións, 2023.

https://doi.org/10.17979/spudc.000024.28

[15] Goel, Siya. Using Feature Selection and Ensemble Algorithms for Sophisticated Pancreatic Cancer Diagnosis. HAL openscience, hal-03119399f, 2021.

[16] Bakasa, Wilson, and Serestina Viriri. Pancreatic Cancer Survival Prediction: A Survey of the State-of-the-Art. Computational and Mathematical

Methods in Medicine, vol. 2021, pp. 1–17, 2021. https://doi.org/10.1155/2021/1188414

[17] Baek, Bin, and Hyunju Lee. Prediction of survival and recurrence in patients with pancreatic cancer by integrating multi-omics data. Scientific reports, vol. 10, no. 1, 2020. https://doi.org/10.1038/s41598-020-76025-1

[18] Keyl, Julius, Stefan Kasper, Marcel Wiesweg, Julian Götze, Martin Schönrock, Marianne Sinn, Aron Berger et al.. Multimodal survival prediction in advanced pancreatic cancer using machine learning. ESMO open 7, no. 5: 100555, 2022. https://doi.org/10.1016/j.esmoop.2022.100555

[19] Bertsimas, Dimitris, Georgios Antonios Margonis, Yifei Huang, Nikolaos Andreatos, Holly Wiberg, Yu Ma, Caitlin Mcintyre et al. Toward an Optimized Staging System for Pancreatic Ductal Adenocarcinoma: A Clinically Interpretable, Artificial Intelligence–Based Model. JCO clinical cancer informatics 5: pp. 1220-1231, 2021. https://doi.org/10.1200/cci.21.00001

[20] The Cancer Imaging Archive https://www.cancerimagingarchive.net/

[21] Krizhevsky, Alex, et al. "ImageNet Classification with Deep Convolutional Neural Networks." Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017. https://doi.org/10.1145/3065386

[22] Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), vol. 9, no. 10, p. p9420, 2019. https://doi.org/10.29322/ijsrp.9.10.2019.p9420

[23] Y. Shi, T. Zhou, Z. You, J. Chu and T. Li. AM-DenseNet: A Novel DenseNet Framework using Attention Mechanisms for COVID-19 CT Image Classification. IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS), Chengdu, China, pp. 474-479, 2022. https://doi.org/10.1109/ccis57298.2022.10016401

[24] T. Kobayashi. SCW-SGD: Stochastically Confidence- Weighted SGD. IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, pp. 1746-1750, 2020. https://doi.org/10.1109/icip40778.2020.9190992

[25] B. Dhananjay, N. P. Venkatesh, A. Bhardwaj and J. Sivaraman. Cardiac signals classification based on Extra Trees model. 8th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 402-406, 2021. https://doi.org/10.1109/spin52536.2021.9565992

[26] V. Vijay and P. Verma. Variants of Naïve Bayes Algorithm for Hate Speech Detection in Text Documents. International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, pp. 18-21, 2023. https://doi.org/10.1109/aisc56616.2023.10085511

[27] M. Kolarik, R. Burget and K. Riha. Comparing Normalization Methods for Limited Batch Size Segmentation Neural Networks. 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, pp. 677-680, 2020. https://doi.org/10.1109/tsp49548.2020.9163397

[28] R. Guedrez, O. Dugeon, S. Lahoud and G. Texier. Label encoding algorithm for MPLS Segment Routing. 2016 IEEE 15th International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, 2016, pp. 113-117, 2016. https://doi.org/10.1109/nca.2016.7778603

[29] N. A. M. Ariff and A. R. Ismail. Study of Adam and Adamax Optimizers on AlexNet Architecture for Voice Biometric Authentication System. 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Republic of Korea, pp. 1-4, 2023. https://doi.org/10.1109/imcom56909.2023.10035592

[30] P. Rasouli and I. C. Yu. Explainable Debugger for Black-box Machine Learning Models. International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, pp. 1-10, 2021. https://doi.org/10.1109/ijcnn52387.2021.9533944

[31] J. Rebane, I. Samsten, P. Pantelidis and P. Papapetrou. Assessing the Clinical Validity of Attention-based and SHAP Temporal Explanations for Adverse Drug Event Predictions. IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, pp. 235-240, 2021. https://doi.org/10.1109/cbms52027.2021.00025

[32] H. Gong, P. Zuliani, Q. Wang and E. M. Clarke. Formal analysis for logical models of pancreatic cancer. 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, USA, pp. 4855-4860, 2011. https://doi.org/10.1109/cdc.2011.6161052