29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Preoperative Features-based Machine Learning Models with XAI for Improved Pancreatic Cancer Survival Prediction

Erica Mae V. Antonino[a], Mikaella Louise D. Layug[a], Giancarlo Gabriel T. Trani[a], Ma Sheila A. Magboo[a], Vincent Peter C. Magboo[a]*

[a]University of the Philippines Manila ,Manila, Philippines

Abstract

Pancreatic cancer is the sixth principal cause of cancer-related deaths with a dismal survival rate largely due to an aggressive tumor behavior, often asymptomatic in the early phase of the disease and already emerging with nodal and distant spread at the time of initial diagnosis. An effective risk stratification for survival needs to be made to determine options for individualized and precise treatment options. Accurate prediction of pancreatic cancer survival remains a daunting challenge due to diverse tumor characteristics and varying clinical presentation at the time of diagnosis. In this study, several machine learning models were examined to predict survival outcomes based on pre-operative clinical and laboratory attributes. Random forest obtained the highest area under the precision recall curve (77.92), excellent recall (96.70) and F2-score (90.16) clearly pointing to its capability to predict reliably the positive class or patients who shall survive for more than one year where surgical resection would be very valuable due to its potential cure. The excellent recall also minimizes the false negatives thereby lessening the likelihood of withholding surgical treatment as potential cure for patients surviving more than one year. The most important attributes to determine survival for more than one year based on SHAP were abdominal pain, CA19-9 (carbohydrate antigen 19-9) and CEA (carcinoembryonic antigen). The adequacy of predictive capacity clearly demonstrates the feasibility of integrating machine learning models into clinical practice which can provide valuable diagnostic and prognostic support to both physicians and patients. This would pave the way for an enhanced decision-making process in identifying patients who shall benefit from curative surgical tumor resection and improving the planning of social and care needs of pancreatic cancer patients leading to a positively enhanced outlook for this dreaded malignancy given its current abysmal prognosis.

---

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: vcmagboo@up.edu.ph

1. Background

Globally, pancreatic cancer is the sixth principal cause of cancer-related deaths with a low overall five-year survival rate of only 10% [1]. This very low survival rate is largely propelled by aggressive tumor behavior, often asymptomatic in the early phase of the disease and already emerging with nodal and distant spread at the time of initial diagnosis [2]. The usual clinical presentation such as epigastric pain extending to the back, nausea, weight loss, fatigue and jaundice, are most often non-specific and contribute heavily to the delay in diagnosis [3]. Risk factors of this malignancy encompass modifiable factors such as obesity, cigarette smoking, diabetes mellitus, chronic pancreatitis, and alcohol consumption, and non-modifiable factors including sex, race, and genetic background [4]. At present, there are no blood markers that depict pancreatic cancer with acceptable accuracy and reliability. In practice, testing for a variety of blood tumor markers such as CEA, CA19-9 etc. in the washing fluid of an endoscopic ultrasound guided fine needle aspiration in combination with other imaging results are being resorted to in order to delineate benign and malignant pancreatic lesion [5]. This seeming lack of specificity renders pancreatic cancer highly prone to underdiagnosis. The imaging modality of choice to assess pancreatic cancer is through contrast-enhanced CT scan due to its potential to assess boundaries of tumor and presence of disease spread [6]. It is the preferred imaging procedure to determine if surgical resection is warranted.

Unfortunately, the diagnosis of pancreatic cancer in its early phase is seen in only a miniscule percentage of patients with the majority already presenting with nodal or distant metastasis at the time of diagnosis. As a result, this has a significant impact on the disease prognosis as treatment of pancreatic cancer is heavily dependent on the surgical resectability of the tumor at the time of diagnosis. As such, very few patients have the chance to be subjected to a surgical resection. Pancreatic tumors farther away from the head of the pancreas will typically require a distal pancreatectomy while those situated in the head and uncinate process are usually subjected to pancreaticoduodenectomy (Whipple surgery) for complete surgical resection [7]. While surgical resection enhances patient's prognosis, offering even potentially a cure if instituted in the early phase of the disease, immediate and long-term complications of the surgical intervention such as weakness, new-onset diabetes, infections, abdominal pain etc. hindering their ability to work or engage in daily activities are very common which strongly influence the patients remaining quality of life [8][9].

Given the worsening outlook for this malignancy, it is imperative that an effective risk stratification for survival as well as recurrence after surgery be made to determine options for individualized and precise treatment modalities for the patients [10]. It is in this area of concern where machine learning (ML) approaches can be harnessed to assess or predict survival using pre-operative attributes. It is crucial to recognize prognostic factors that guide in the determination of good surgical candidates and restrict patients with an expected dismal end result from being subjected to a major surgery [9]. Very few studies have been published regarding the assessment of survival in pancreatic cancer using pre-operative features with added interpretability capability. To address this research gap, the goal of the study is to predict survival outcomes of pancreatic cancer with model interpretability based on a mixture of pre-operative clinical and laboratory attributes utilizing a variety of ML models namely: logistic regression (LR), decision tree (DT), Naïve Bayes (NB), support vector machine (SVM), k-Nearest Neighbors (kNN), extra trees (ET), Bootstrap aggregating (BG), extreme gradient boosting (XGB), gradient boosting (GB) and random forest (RF).

There has been an increasing trend taking advantage of these artificial intelligence methods to guide in medical decision making for many types of diseases. The significant contribution of this study pertains to the development of automated tools in the precise and reliable prediction of survival and disease progression necessary for improving diagnostic and therapeutic management options for the patients including specifying potential eligibility for clinical trials. These predictive models aided by artificial intelligence are essential for recognizing patients at higher risk of rapid disease worsening, enabling personalized treatment plans. As such, ML models could serve as valuable decision support tools for both patients and physicians where surgical intervention could be offered after an assessment of the patient's survival chance.

The outline of this work commences with a backgrounder section followed by a survey of relevant literature focusing on ML approaches in predicting survival for pancreatic cancer. Section 3 pertains to the methodology

detailing the dataset characteristics, preprocessing steps and machine learning configurations and performance indices. Subsequently, Section 4 highlights the main findings together with a comprehensive discussion. The last section provides the main summary and recommendations for future research endeavors.

2. Survey of Literature

The utility of artificial intelligence (AI) in the clinical setting has been the focus of many published studies. In a review by Tripathi et al., AI has demonstrated its impact in personalized medicine by evaluating patient clinical data to suit therapeutic options in pancreatic cancer [11]. The review has shown the potential of AI in reshaping cancer care through enhanced and more precise diagnostics, personalized therapeutic options paving the way for improved clinical outcomes. In [12], authors likewise reported the clinical usefulness of AI in pancreatic cancer care as it has the capability to predict duration of survival, possibility of recurrence, assessing response to therapy by analyzing medical imaging, tumor markers, pathological examination datasets. In the study by Teng et al., authors have evaluated a variety of ML algorithms to predict survival of pancreatic cancer using the Surveillance, Epidemiology, and End Results (SEER) dataset [13]. Results showed CatBoost to be the top performing ML model with an area under the curve (AUC) of 0.826, 0.839 accuracy, 0.872 recall and 0.832 precision. Likewise, Shen et al., employed several ML models to predict a 12-month survivorship in pancreatic cancer patients with kNN classifier obtaining the highest AUC of 86.1% [14]. Authors further reported that that most important predictors included C-Reactive protein (CRP), surgical intervention, carbohydrate antigen (CA)19-9, age at diagnosis, and the neutrophil–lymphocyte ratio (NLR) and further concluded that ML models to be useful for improved risk stratification leading to individualized prognostic appraisal in this patient cohort.

In [9], authors analyzed a group of 93 patients to determine prognostic features in resectable pancreatic head adenocarcinoma and further developed a ML algorithm to assess survival for patients who were subjected to a pancreaticoduodenectomy. The prognostic features included type 2 diabetes, family history of cancer, perineural involvement, bile duct stricture, portal vein tissue resection, neoadjuvant therapy, tumor size based on MRI, and tumor size based on histopathology. The ML algorithm yielded a 75% accuracy, 41.9% sensitivity, and 97.5% specificity in predicting survival for less than 2 years after surgery. Similarly, Vigia et al. examined preoperative features bringing about discouraging 1-year survival prognosis in patients with pancreatic ductal adenocarcinoma [15]. These included gender, age, tumor marker Ca-19-9, jaundice, ERCP with biliary stent, neutrophils, lymphocytes, lymphocyte/neutrophil ratio, neoadjuvant treatment, imaging tumor size, and ASA. Authors also employed a logistic regression model to assess a 1-year survival prognosis with an AUC of 0.74 and accuracy of 0.78 which can be utilized as a decision support tool to an informed decision-making in the evaluating which patients stand to gain from undergoing surgical procedure (pancreatoduodenectomy). In [16], authors examined an ensemble machine learning combining clinical data and computed tomography (CT) data to determine survival preoperatively generating an AUC of 0.76 for predicting 2-year overall survival. Authors concluded that the ensemble model generated a similar capability in predicting 2-year overall survival as compared to the American Joint Committee on Cancer.

Huang et al. analyzed long-term survival and identifying important prognostic factors using ML among patients with borderline resectable pancreatic cancer [17]. Authors have enumerated these prognostic factors for long-term survival as age, length of vascular invasion vascular morphological malformation, and regional lymph nodes. The logistic regression model has predicted long-term survival with an AUC 0.875 and has concluded that the model can provide decision support for surgeons in determining surgical intervention options for pancreatic cancer patients. In the study by Ahmed et al., authors developed an ML model to forecast 1-year survival of patients subjected to pancreatectomy for pancreatic cancer using the American Society of Clinical Oncology (ASCO)'s CancerLinQ Discovery (CLQD) [18]. Results showed random forest and logistic regression predicted the 1-year survival with an AUC of 0.83 each and identified the following to be the most important predictors for survival to be: chemotherapy, tumoral size, pre-operative body mass index, age, pre-operative neutrophil count, serum cholesterol, and serum CA-19-9 levels. Authors concluded that ML models can provide reliable and precise prediction of 1-year survival using clinicopathologic, demographic, and laboratory data provide and as such, these models can be helpful in determining patients at heightened risk for dismal outcomes after pancreatic surgery for which alternative therapeutic options may be warranted.

In summary, many of the published studies have focused on early detection of pancreatic cancer through artificial intelligence while fewer studies have addressed survival prediction, particularly using preoperative clinical feature datasets. Moreover, the role of methodological enhancements such as feature selection techniques and handling class imbalance to improve model performance in survival prediction have not been thoroughly explored. Adding these factors could significantly advance medical research, providing more precise, efficient, and personalized healthcare solutions.

3. Materials and Methods

The research pipeline in this study comprised several phases as seen in Fig. 1. The dataset was extracted from the IEEE DataPort platform. Afterwards, several preparatory stages were employed to ensure that the data was clean, consistent, and adequately prepared for subsequent machine learning tasks. The preprocessed dataset was separated into a training and test set and incorporated feature engineering and feature selection. After model training with several models with hyperparameter tuning, model evaluation ensued with incorporation of SHAP to aid in the interpretability.

Fig. 1. Research Pipeline for Pancreatic Cancer Survival Study.
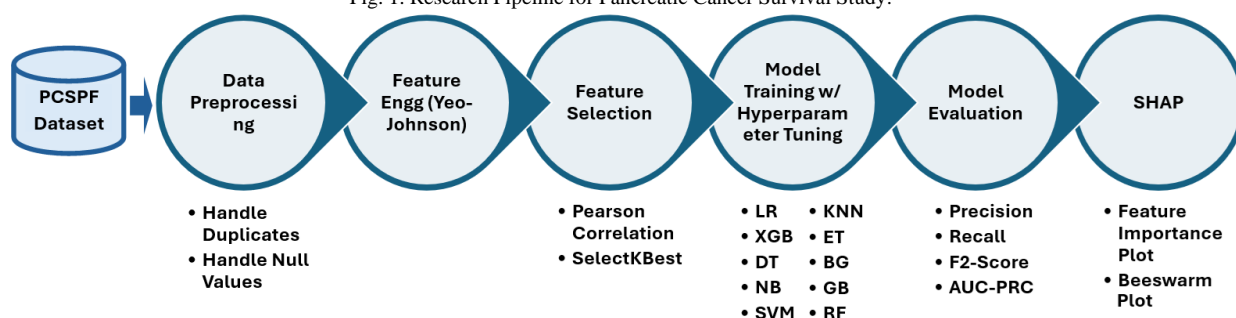
3.1. Dataset Characterization with Preprocessing Steps

The study utilized the PCSPF-Pancreatic Cancer Survival Based on Preoperative Features dataset from the IEEE DataPort [19]. This dataset comprised 878 instances with 20 preoperative features, including both numerical and categorical variables. The numerical variables included age, body mass index (BMI), C-reactive protein (CRP), albumin (ALB), CRP/ALB, leukocyte, neutrocyte, platelet, lymphocyte, neutrocyte lymphocyte ratio (NLR), platelet lymphocyte ratio (PLR), systemic immune-inflammation index (SII), lactic dehydrogenase, carbohydrate antigen 19-9 (CA19-9), carcinoembryonic antigen (CEA), prealbumin, total bilirubin, and directed bilirubin. The categorical variables consisted of sex and abdominal pain. Additionally, there is a target variable that represents the survival outcome with a value of 1 (denoting patients surviving one year or more) and 0 (denoting patients surviving for less than one year). Survival for at least one year is a paramount marker for the outcome of pancreatic cancer patients. Furthermore, the dataset exhibited a mild class imbalance, with 604 patients (68.8%) surviving for one year or longer, and 274 patients (31.2%) surviving for less than one year. As the dataset has already been standardized, the data cleaning only involved checking for duplicates and null values. This showed that no data needed to be removed

3.2. Feature Engineering

Exploratory data analysis was performed prior to modelling using machine learning. The histograms of numerical variables namely: CRP, CRP/ALB, leukocyte, neurocyte, platelet, lymphocyte, NLR, PLR, SII, dehydrogenase, CEA, total bilirubin and directed bilirubin showed positive skewness. As such, the Yeo-Johnson transformation was applied to these features to reduce the heteroscedasticity of these features as well as to boost the normality, thus evolving into a probability density function that is similar to a normal distribution. Additionally, instead of removing or replacing outliers, this transfiguration keeps them in the dataset, which enhances the data structure and lessens the outliers' impact on prediction outcomes [20].

3.3. Feature selection

The heatmap highlighting correlation between variables is shown in Fig. 2. To mitigate the potential adverse effects of multicollinearity on model stability and interpretation, CRP/ALB and directed bilirubin were dropped from the analysis. In this study, a univariate feature selection method (SelectKBest using Analysis of Variance) was then utilized which led to the selection of 10 optimal features. These were the following: abdominal pain, age, CRP, neutrocyte, NLR, SII, lactic dehydrogenase, CA19-9, prealbumin and CEA. The main benefit of employing a feature selection procedure is to dampen potential over-fitting as simpler data is produced for easy interpretability by eliminating features will less prediction importance. Univariate feature selection methods have been shown to be useful when dealing with fewer features as well as being computationally inexpensive, thus easy to implement [21][22].
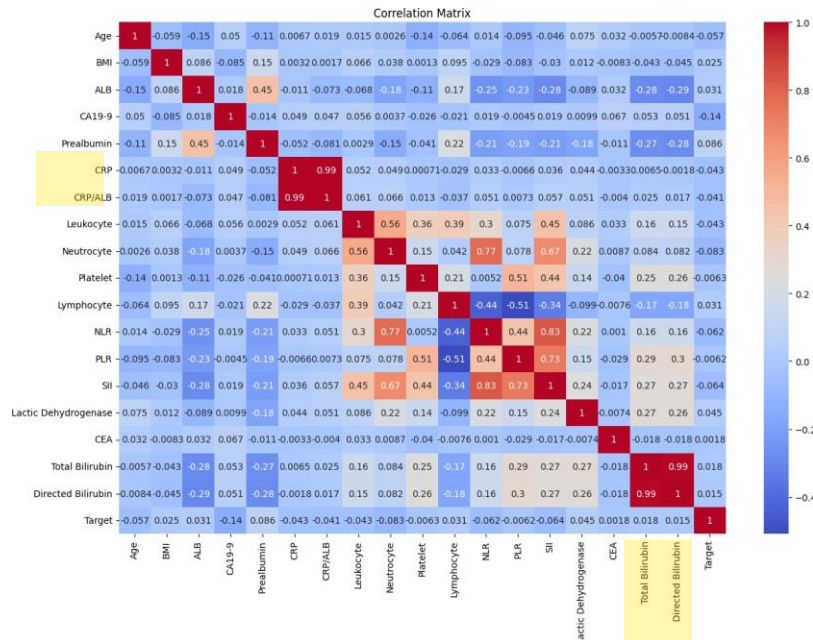

Fig. 2. Correlation heatmap of numerical attributes

3.4. Machine Learning Models

The dataset was subdivided into 70% training and 30% testing sets. To predict pancreatic cancer survival based on preoperative features, various supervised ML models were used in this research work namely: LR, DT, NB, SVM, kNN, ET, BG, XGB, GB and RF. Python 3.10.12 and its ML libraries (scikit-learn, pandas, Matplotlib, seaborn, and NumPy) were utilized in running and evaluating the models.

Hyperparameter tuning was also used to maximize the predictive capacity of the ML models by adjusting the model parameters. Grid Search with Cross Validation was chosen which evaluates the performance of the models achieved through each combination of all possible parameters.

3.1. Model Evaluation and Interpretability

To assess model performance, various indices were calculated, including Precision, Recall, F2-score, and area under the precision recall curve (AUC-PRC). Recall or sensitivity represents how many of those truly will survive for more than one year (true positive and false negative) can be identified correctly by the model (true positive). Precision represents the proportion of those correctly predicted to be surviving for more than one year (true positive) among all those classified by the model as surviving for more than one year (true positive and false positive). The F2-score or (F-Beta Score) is a measure that evaluates the accuracy of an output of a model from two aspects of

Table 1. List of Hyperparameters of the ML Models

ML Model	Hyperparameters
LR	C=0.1
XGB	learning_rate=0.01, max_depth=3, n_estimators=200
DT	max_depth=10, min_samples_split=5
SVM	C=0.1, gamma=0.1
kNN	n_neighbors=10
ET	max_depth=10, min_samples_split=10, n_estimators=300
BG	max_samples=0.7, n_estimators=100
GB	learning_rate=0.01, max_depth=3, n_estimators=100
RF	max_depth=10, min_samples_split=2, n_estimators=300

precision and recall, giving more importance to recall. In this domain, it is believed that recall should be given more weight as compared to precision as it is the intent of the study is to identify those surviving for more than one year (positive class). Generally, this group of patients surviving for more than one year would benefit more from the contemplated surgical intervention as it potentially offers a cure. Likewise, giving more preference to recall would induce minimizing false negatives and would be more detrimental as this is tantamount to withholding potential cure via surgery to these patients surviving for more than one year. A high recall indicates that the model is adequate at recognizing all positive instances, lessening the chance of missing true positives. In this study, the AUC-PRC was chosen to determine the best performing model as the positive class is of higher clinical importance. It is also a particularly useful metric in cases of class imbalance. SHapley Additive exPlanations (SHAP) values were harnessed to aid in the interpretation of the top performing model predictions. SHAP values illustrate how individual attributes influence each ultimate prediction, the importance of each attribute compared to others, and the model's dependency on the interaction between attributes. Attributes having positive SHAP values positively impact the prediction, while those having negative values have a negative impact [23]. Additionally, SHAP values are model-agnostic and as such can be harnessed to aid in the interpretation of any machine learning model [24].

4. Findings and Analysis

The performance indices for predicting survival for more than one year among patients with pancreatic cancer are depicted in Table 2. The top performing model was generated by random forest with the highest AUC-PRC of 77.92, 96.70 recall, 70.97 precision, and 90.16 F2-score. Following very closely were k-nearest neighbors and logistic regression with the AUC-PRC of 77.48 and 77.42 respectively. It is noteworthy to mention that apart from DT and NB, the rest of the models obtained acceptably superior recall or sensitivity ranging from 91.76 to 1.00 as well as F2-scores ranging from 86.62 to 91.92. These strongly suggest the capability of the models to predict reliably the positive class or patients who shall survive for more than one year where surgical resection would be very valuable due to its potential cure. Additionally, the excellent recall also minimizes the false negatives thereby lessening the likelihood of thwarting surgical treatment as potential cure for patients surviving more than one year. The adequacy of predictive capacity clearly demonstrates the feasibility of incorporating these models in actual clinical scenario providing decision support to both physicians and patients, thus paving the way for enhanced decision-making process in identifying patients who shall benefit from surgical tumor resection. With proper guidance from the health professionals, integrating these models into actual clinical setting would lead to a positively enhanced outlook for this dreaded malignancy given its current abysmal prognosis.

The global feature importance rankings based on the mean SHAP value using the optimal random forest model are shown on Fig. 3 while the beeswarm plot is illustrated in Fig. 4. The topmost impactful attributes include abdominal pain, CA19-9 (carbohydrate antigen 19-9) and CEA (carcinoembryonic antigen). Based on Fig. 4, higher counts of abdominal pain help to increase the prediction of not surviving for more than one year, i.e., negatively affecting the output. Additionally, lower values of CA19-9 and CEA tend to positively affect the output leading towards prediction of surviving for more than one year. Similarly, higher values of prealbumin help to increase the prediction of

Table 2. Performance indices for predicting survival more than one year

ML Model	Precision	Recall	F2-score	AUC-PRC
LR	70.73	95.60	89.32	77.42
XGB	69.20	100.00	91.83	76.35
DT	71.73	75.27	74.54	75.22
NB	71.05	89.01	84.73	74.77
SVM	69.11	98.35	90.68	72.92
kNN	70.76	91.76	86.62	77.48
ET	69.65	98.35	90.86	75.48
BG	71.49	95.05	89.18	75.05
GB	69.47	100.00	91.92	75.88
RF	70.97	96.70	90.16	77.92

surviving for more than one year which suggests positively affecting the output. Interestingly, for age and neutrocytes, which may be lesser important attributes globally as shown in Fig. 4, higher values of age and neutrocytes however, seem to be predictive of not surviving for more than one year for some patients as depicted in Fig. 5.

While abdominal pain, a non-specific symptom, is a common presenting manifestation of pancreatic cancer, it has not been reported to be directly related to the risk of survival for more than one year. However, the presence of severe or persistent abdominal pain may denote an advanced phase of malignancy with the severity of pain directly associated with worsening prognosis. Perineural invasion is a notable clinical presentation of pancreatic cancer which usually suggests being the route of tumoral spread (metastasis) producing neuropathic pain [25]. This tumoral spread or nerve involvement can indirectly influence survival rate and thus negatively influence disease prognosis. On the other hand, elevated values of biomarkers CEA and CA19-9 have been reported to be significant barometers of a malignant phenotype, particularly a surgically unresectable tumor or tumor being more aggressive, with lymph nodes that cannot be adequately assessed, advanced stages and distant metastases. Thus, elevated serum levels of CEA and CA19-9 strongly suggest lower overall survival rates warranting more watchful follow-up [26]. Likewise, higher values of CA19-9 strongly points to lower survival in pancreatic cancer patients while longer survival can be seen in the setting of a decreased CA19-9 levels after treatment. Elevated CA19-9 levels may signal the presence of micro-metastases such as hepatic microspread which can possibly remain unnoticed at the time of surgery [27]. Additionally, age is a critical factor in the evaluation of survival rates as generally, younger patients tend to have longer survival rates presumably brought about by better overall health and fewer comorbidities as compared to older patients who may have other pre-existing comorbidities. Higher neutrophil counts have been linked to dismal survival rate in patients with pancreatic cancer particularly after adjuvant chemotherapy. Elevated neutrophils, being an inflammatory marker of malignancy, have been linked to atrocious outcomes of overall survival, progression-free survival and recurrence-free survival and thus require additional therapeutic options in advanced disease [28].

The results of this study were similar to other published studies on the survival of patients with pancreatic cancer, particularly that of Lee et al. [16] and Vigia et al. [15]. As compared to [13], [14] and [18], this study has a slightly lower predictive capacity but has a significantly higher recall as compared to that of Baig et al. [9].

The predominant limitation of this research study lies in the choice of the dataset as it has already been standardized. The incorporation of various feature selection and techniques will be limited as it might reduce the data integrity. It is therefore warranted to utilize raw data samples to have an accurate understanding of features and their impact on pancreatic cancer patient survival. Due to the small size of the tabular dataset, the utilization of deep learning algorithms was not feasible; however, these algorithms could be advantageous in image analysis, particularly for identifying and segmenting pancreatic cancer in medical images like CT scans or MRI scans.
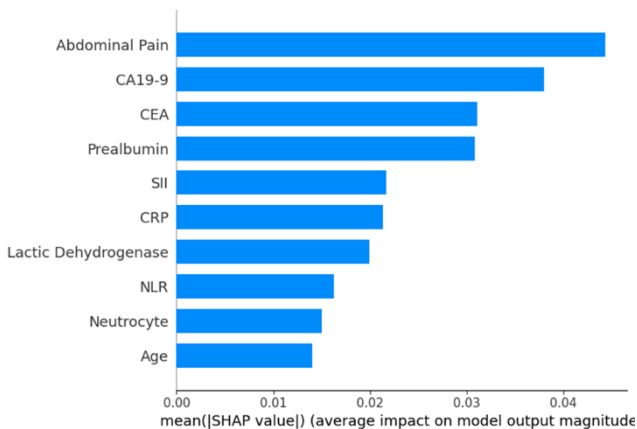
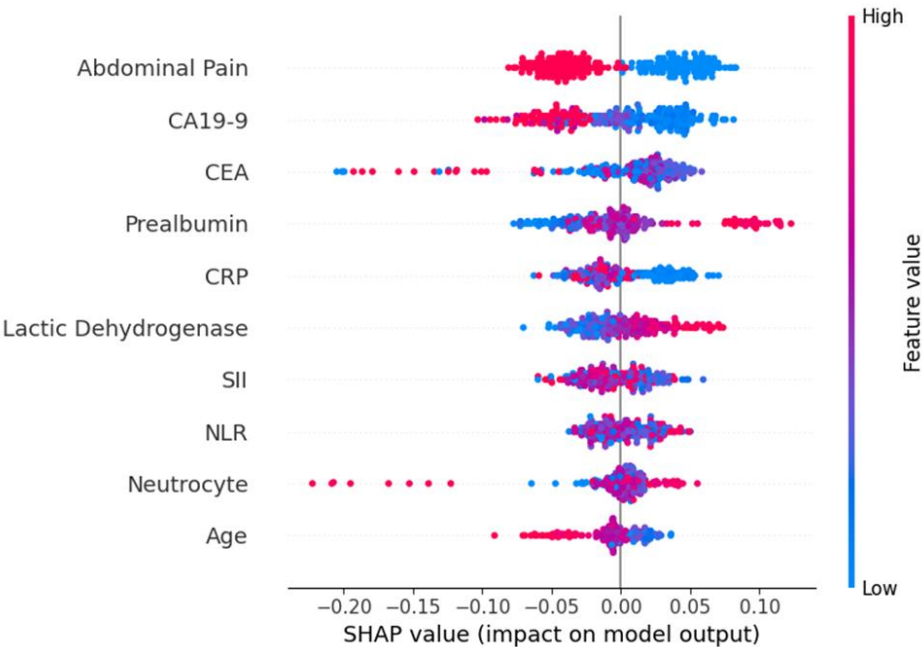Fig. 3. Feature Importance Ranking for Random Forest based on Mean SHAP value

Fig. 4. Beeswarm Plot for Random Forest

5. Synopsis and Recommendations

Pancreatic cancer is the sixth principal cause of cancer-related deaths with a dismal rate of survival largely due to an aggressive tumor behavior, often asymptomatic in the early phase of the disease and already emerging with nodal and distant spread at the time of initial clinical evaluation. This has a tremendous impact on the disease prognosis as treatment of pancreatic cancer is heavily dependent on the surgical removal of the tumor at the time of diagnosis. An effective risk stratification for survival needs to be made to determine options for individualized and precise treatment modalities for the patients. It is in this area of concern where machine learning approaches can be harnessed to assess or predict the survival of pancreatic cancer. In this study, various machine learning models were analyzed to predict

survival outcomes based on pre-operative clinical and laboratory attributes with random forest generating the highest area under the precision recall curve, excellent recall and F2-score which indicate its capacity to predict reliably the positive class or patients who shall survive for more than one year where surgical resection would be very valuable due to its potential cure. Additionally, the excellent recall also minimizes the false negatives thereby lessening the likelihood of thwarting surgical treatment as potential cure for patients surviving more than one year. The most important attributes to determine survival for more than one year based on SHAP were abdominal pain, CA19-9 (carbohydrate antigen 19-9) and CEA (carcinoembryonic antigen). SHAP can be utilized to aid in the interpretation of the best performing model (random forest).

The adequacy of predictive capacity clearly demonstrates the feasibility of incorporating these models in actual clinical scenario providing decision support to both physicians and patients, thus paving the way for enhanced decision-making process in identifying patients who shall benefit from surgical tumor resection. With proper guidance from the health professionals, integrating these models into actual clinical setting would lead to a positively enhanced outlook for this dreaded malignancy given its current abysmal prognosis.

References

[1] Leiphrakpam, Premila Devi, Sanjib Chowdhury, Michelle Zhang, Varnica Bajaj, Mashall Dhir, and Chandrakanth Are. (2025). "Trends in the Global Incidence of Pancreatic Cancer and a Brief Review of its Histologic and Molecular Subtypes." Journal of gastrointestinal cancer, 56(1), 71. https://doi.org/10.1007/s12029-025-01183-2.

[2] Stoop, Thomas, Ammar Javed, Atsushi Oba, et al. (2025) "Pancreatic cancer." The Lancet, 405 (10485), 1182-1202. https://doi.org/10.1016/S0140-6736(25)00261-2.

[3] Grigorescu, Raluca Roxana, Iona Alexandra Husar-Sburlan, and Cristian Gheorghe. (2024). "Pancreatic Cancer: A Review of Risk Factors." Life, 14(8), 980. https://doi.org/10.3390/life14080980

[4] Yu, Weidong, Danyi Zhou, Fanhao Meng, et al. (2025) "The global, regional burden of pancreatic cancer and its attributable risk factors from 1990 to 2021." BMC Cancer 25, 186. https://doi.org/10.1186/s12885-025-13471-y.

[5] Li, Chen-Fan, Lan-Xiang Hao, Yi-Jia Chen, et al. (2025). "A Novel Diagnostic Model of Biomarkers in the Washing Fluid Obtained by EUS-FNA in Pancreatic Cancer." Journal of gastrointestinal cancer, 56(1), 89. https://doi.org/10.1007/s12029-025-01209-9.

[6] Anghel, Cristian, Mugur Cristian Grasu, Denisa Andreea Anghel, Gina-Ionela Rusu-Munteanu, Radu Lucian Dumitru, and Ioana Gabriela Lupescu. (2024) "Pancreatic Adenocarcinoma: Imaging Modalities and the Role of Artificial Intelligence in Analyzing CT and MRI Images." Diagnostics, 14(4), 438. https://doi.org/10.3390/diagnostics14040438.

[7] Hayat, Umar, Phillip S. Croce, Aseel Saadeh, et al. (2025) "Current and Emerging Treatment Options for Pancreatic Cancer: A Comprehensive Review." Journal of Clinical Medicine, 14(4), 1129. https://doi.org/10.3390/jcm14041129.

[8] Chauhan, Priyanka, Sukhpal Kaur, Neena Vir Singh, Maninderdeep Kaur, and Rajesh Gupta, (2025) "Lived Experiences of Pancreatic Cancer Patients Undergone Whipple Procedure: A Qualitative Study." Indian journal of palliative care, 31(1), 67–73. https://doi.org/10.25259/IJPC_205_2024.

[9] Baig, Zarrukh, Nawaf Abu-Omar, Yigang Luo, Rayyan Khan, et al. (2021) "Prognosticating Outcome in Pancreatic Head Cancer With the use of a Machine Learning Algorithm." Technology in Cancer Research & Treatment. 2021;20. doi:10.1177/15330338211050767.

[10] Qu, Chao, Piaoe Zeng, Changlei Li, et al. (2025) "A machine learning model based on preoperative multiparametric quantitative DWI can effectively predict the survival and recurrence risk of pancreatic ductal adenocarcinoma." Insights Imaging 16, 38. https://doi.org/10.1186/s13244-025-01915-9.

[11] Tripathi, Sathvik, Azadeh Tabari, Arian Mansur, Harika Dabbara, Christopher P Bridge, and Dania Daye. (2024). "From Machine Learning to Patient Outcomes: A Comprehensive Review of AI in Pancreatic Cancer." Diagnostics, 14(2), 174. https://doi.org/10.3390/diagnostics14020174.

[12] Ozsahin, Dilber Uzun, Natacha Usanase and Ilker Ozsahin Ozsahin. (2025) "Advancing pancreatic cancer management: the role of artificial intelligence in diagnosis and therapy". Beni-Suef Univ J Basic Appl Sci 14, 32 (2025). https://doi.org/10.1186/s43088-025-00610-4.

[13]Teng, Buwei, Xiaofeng Zhang, Mingshu Ge, Miao Miao, Wei Li, and Jun Ma. (2024) "Personalized three-year survival prediction and prognosis forecast by interpretable machine learning for pancreatic cancer patients: a population-based study and an external validation." Frontiers in Oncology 14. doi: 10.3389/fonc.2024.1488118.

[14] Shen, Yan-Chen, Po-See Chen, Cheng-Feng Lin, et al. (2024) "Predictive machine learning models for survival outcomes in patients with pancreatic cancer." Journal of Clinical Oncology 42: 625-625. doi:10.1200/JCO.2024.42.3_suppl.625

[15] Vigia, Emmanuel, Luis Ramalhete, Edite Filipe, et al. (2023) "Machine Learning-Based Model Helps to Decide which Patients May Benefit from Pancreatoduodenectomy." Onco, 3(3), 175-188. https://doi.org/10.3390/onco3030013.

[16] Lee, Woohyung, Hyo Jung Park, Hack-Jin Lee, et al. (2022) "Preoperative data-based deep learning model for predicting postoperative survival in pancreatic cancer patients." International Journal of Surgery, 105, 106851. https://doi.org/10.1016/j.ijsu.2022.106851.

[17] Huang, Jin-Can, Shao-Cheng Lyu, Bing Pan, et al. (2025). "A logistic regression model to predict long-term survival for borderline resectable pancreatic cancer patients with upfront surgery." Cancer imaging: the official publication of the International Cancer Imaging Society, 25(1), 10. https://doi.org/10.1186/s40644-025-00830-y.

[18] Ahmed, Kaleem Sohail, Sheriff M Issaka, Clayton Thomas Marcinak, and Syed Nabeel Zafar. (2024). "Predicting real-world overall survival for post-operative pancreatic ductal adenocarcinoma: An ASCO CancerLinQ Discovery analysis." JCO 42, 4169-4169. DOI:10.1200/JCO.2024.42.16_suppl.4169

[19] Beilei Wang. (2023). PCSPF-Pancreatic Cancer Survival based on Preoperative Features. IEEE Dataport. https://dx.doi.org/10.21227/gj12-n882.

[20] Sun, Lijun, Nanyan Hu, Yicheng Ye, et al. (2022) "Ensemble stacking rockburst prediction model based on Yeo–Johnson, K-means SMOTE, and optimal rockburst feature dimension determination." Sci Rep 12, 15352. doi: 10.1038/s41598-022-19669-5

[21] Abellana, Dharyll Prince and Demelo M. Lao. (2023) "A new univariate feature selection algorithm based on the best–worst multi-attribute decision-making method." Decision Analytics Journal, 7, 100240. https://doi.org/10.1016/j.dajour.2023.100240.

[22] Islam, Mainul and Rafiqul Islam. (2023) "Exploring the Impact of Univariate Feature Selection Method on Machine Learning Algorithms for Heart Disease Prediction." 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, pp. 1-5. doi: 10.1109/NCIM59001.2023.10212832.

[23] Lamane, Houda, Latifa Mouhir, Rachid Moussadek, et al. (2025). "Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration." International Journal of Sediment Research, 40 (1), 91-107. https://doi.org/10.1016/j.ijsrc.2024.10.002.

[24] Wang, Huanjing, Quanxin Liang, John T Hancock, and Taghi M. Khoshgoftaar. (2024) "Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods." J Big Data 11, 44. https://doi.org/10.1186/s40537-024-00905-w.

[25] Lohse, Ines, and Shaun P Brothers. (2020) "Pathogenesis and Treatment of Pancreatic Cancer Related Pain." Anticancer research, 40(4), 1789–1796. https://doi.org/10.21873/anticanres.14133.

[26] Ermiah, Eramah, Mona Eddfair, Othman Abdulrahman, et al. (2022) "Prognostic value of serum CEA and CA19-9 levels in pancreatic ductal adenocarcinoma." Molecular and Clinical Oncology, 17, 126. https://doi.org/10.3892/mco.2022.2559.

[27] Kang, Yong Ming, Hao Wang, Ran Li, and Gu Pan. (2021) "Prognostic Role of Carbohydrate Antigen 19 to 9 in Predicting Survival of Patients With Pancreatic Cancer: A Meta-Analysis." Technology in Cancer Research & Treatment. 2021;20. doi:10.1177/15330338211043030.

[28] Worapongpaiboon, Rinrada, Noppachai Siranart, Patavee Pajareya, Somkiat Phutinart. (2025). "Inflammatory markers in predicting survival in pancreatic cancer: A Systematic review and Meta-Analysis." Pancreatology, 2025. https://doi.org/10.1016/j.pan.2025.02.014.