İnönü Üniversitesi Sağlık Hizmetleri Meslek Yüksekokulu Dergisi
ISSN: 2147-7892, Cilt 13, Sayı 1 (2025) 62-73
doi: 10.33715/inonusaglik.1571883

Journal of Inonu University Health Services Vocational School
ISSN: 2147-7892, Volume 13, Issue 1 (2025) 62-73
doi: 10.33715/inonusaglik.1571883

Original Article / Araştırma Makalesi

IDENTIFICATION AND GLOBAL INTERPRETATION OF POSSIBLE BIOMARKERS FOR THE DIAGNOSIS OF PANCREATIC CANCER USING EXPLAINABLE ARTIFICIAL INTELLIGENCE METHODS

Açıklanabilir Yapay Zekâ Yöntemleri Kullanılarak Pankreas Kanseri Tanısı için Olası Biyobelirteçlerin Belirlenmesi ve Global Olarak Yorumlanması

Şeyma YAŞAR[1]
[1]*İnönü Üniversitesi, Tıp Fakültesi, Malatya*

Geliş Tarihi / Received: *23.10.2024* ***Kabul Tarihi / Accepted:*** *20.01.2025*

ABSTRACT

Pancreatic cancer is a highly lethal malignancy with poor prognosis and limited early diagnosis methods. In this study, 60 serum samples (30 pancreatic cancer patients, 30 controls) were analyzed to identify potential biomarkers for early detection using machine learning. Proteomic data were obtained via glycoprotein enrichment and mass spectrometry, identifying 232 proteins. After preprocessing, 29 proteins were selected using the Elastic Net method. XGBoost, optimized with 10-fold cross-validation, classified pancreatic cancer with high performance (AUC=0.850, accuracy=0.833). The SHAP method identified P02750 (Leucine-rich alpha-2-glycoprotein), P02766 (Transthyretin), P01031 (Complement C5), and P02649 (Apolipoprotein E) as key proteins affecting cancer risk. These biomarkers may play a crucial role in early diagnosis and personalized treatment, but further validation in larger studies is required.

Keywords: Biomarker, Explainable artificial intelligence, Pancreatic cancer, XGBoost.

ÖZ

Pankreas kanseri, kötü prognozu ve sınırlı erken tanı yöntemleri ile oldukça ölümcül bir malignitedir. Bu çalışmada, 60 serum örneği (30 pankreas kanseri hastası, 30 kontrol) makine öğrenimi kullanılarak erken teşhis için potansiyel biyobelirteçleri belirlemek üzere analiz edilmiştir. Proteomik veriler glikoprotein zenginleştirme ve kütle spektrometrisi yoluyla elde edilmiş ve 232 protein tanımlanmıştır. Ön işlemeden sonra, 29 protein Elastik Ağ yöntemi kullanılarak seçilmiştir. XGBoost, 10 kat çapraz doğrulama ile optimize edilmiş, pankreas kanserini yüksek performansla sınıflandırmıştır (AUC= 0.850, doğruluk = 0.833). SHAP yöntemi, P02750 (Lösinden zengin alfa-2-glikoprotein), P02766 (Transtiretin), P01031 (Complement C5) ve P02649'u (Apolipoprotein E) kanser riskini etkileyen anahtar proteinler olarak tanımlamıştır. Bu biyobelirteçler erken tanı ve kişiselleştirilmiş tedavide önemli bir rol oynayabilir, ancak daha büyük çalışmalarda daha fazla doğrulama yapılması gerekmektedir.

Anahtar kelimeler: Açıklanabilir yapay zekâ, Biyobelirteç, Pankreas kanseri, XGBoost.

Şeyma YAŞAR ✉, seyma.yasar@inonu.edu.tr
İnönü Üniversitesi, Tıp Fakültesi, Malatya

Bu makaleye atıf yapmak için (How to cite this article): Yaşar, Ş. (2025). Identification and Global Interpretation of Possible Biomarkers for The Diagnosis of Pancreatic Cancer Using Explainable Artificial Intelligence Methods. *İnönü Üniversitesi Sağlık Hizmetleri Meslek Yüksekokulu Dergisi*, 13(1), 62-73. doi: 10.33715/inonusaglik.1571883

INTRODUCTION

Pancreatic cancer is a type of malignancy caused by the uncontrolled proliferation of cells in pancreatic tissue and has an extremely poor prognosis. The pancreas is an organ with both endocrine and exocrine functions that play critical roles in digestion and metabolism. Pancreatic cancer is difficult to diagnose at an early stage as it usually manifests itself in advanced stages, and by the time most patients are diagnosed, the disease has reached the metastatic stage. For this reason, pancreatic cancer is considered one of the deadliest cancers worldwide, with a five-year survival rate of less than 10%. These challenges in diagnosis and treatment increase the need for new methods to better understand and manage pancreatic cancer. Approximately 495,000 new cases of pancreatic cancer are diagnosed worldwide each year. The incidence of cancer varies depending on geographical regions and lifestyles, with higher rates in developed countries. Pancreatic cancer, which is more common in men than women, is mostly diagnosed in people aged 65 and over. Genetic and environmental factors play a role in the development of pancreatic cancer. Smoking is recognized as one of the most important environmental risk factors that increase the risk of pancreatic cancer 2-3 times. In addition, obesity and physical inactivity increase the risk of pancreatic cancer development. Factors such as a family history of pancreatic cancer, genetic mutations such as BRCA2 and PALB2, and long-term diabetes also increase the risk (Ansari, Torén, Zhou, Hu, & Andersson, 2019; Mizrahi, Surana, Valle, & Shroff, 2020). Diabetes is considered both a risk factor and one of the early symptoms of pancreatic cancer. Long-term inflammation of the pancreas, such as chronic pancreatitis, also increases susceptibility to cancer. Traditional methods used in the diagnosis of pancreatic cancer include medical history, physical examination, laboratory tests and imaging methods (ultrasonography, computed tomography and magnetic resonance imaging). Tissue samples taken by biopsy provide confirmation of the diagnosis through pathologic examinations. Only cases diagnosed at an early stage with a resectable tumor are eligible for surgical treatment, one of the most effective treatment methods for early-stage pancreatic cancer. Regrettably, approximately 80% of patients receive a diagnosis at an advanced stage, thereby forfeiting their opportunity for surgical treatment. Patients diagnosed at an advanced stage receive systemic approaches such as chemotherapy, radiotherapy, and targeted therapies (Hanna-Sawires et al., 2021). Chemotherapy options, on the other hand, have limited effectiveness and can cause serious side effects. Despite the use of folfırınox or gemcitabine-based treatment protocols, the development of tumor cell resistance to these drugs poses a significant problem (Halbrook, Lyssiotis, di Magliano, & Maitra, 2023; Kolbeinsson, Chandana, Wright, & Chung, 2023). At

this point, the identification of proteomic biomarkers is of critical importance. Thanks to new-generation proteomic technologies, it is possible to identify biomarkers that can provide early diagnosis of pancreatic cancer, predict the course of the disease, and predict the response to treatment. Protein-based biomarkers, particularly those obtained through liquid biopsy, enable minimally invasive disease monitoring.

The role of biomarkers in pancreatic cancer is becoming increasingly important in the diagnosis and treatment processes. In particular, CA 19-9 biomarker is a widely used marker in the diagnosis of pancreatic cancer. However, the diagnostic sensitivity and specificity of biomarkers may be limited in this cancer type. Therefore, new approaches are needed to identify more effective biomarkers and to better understand the molecular basis of the disease. In recent years, artificial intelligence (AI) and explainable artificial intelligence (XAI) methods have revolutionized the diagnosis and treatment of complex diseases such as pancreatic cancer. By analyzing large biological data sets, AI offers great advantages in identifying disease risk factors, improving diagnostic accuracy and predicting treatment outcomes. Especially in heterogeneous and multifactorial diseases such as pancreatic cancer, AI allows for more precise identification of biomarkers and a better understanding of cancer subtypes at the molecular level. The machine learning algorithms such as Random Forest, XGBoost, Support Vector Machines and LightGBM have been widely used to identify these biomarkers. Explainable artificial intelligence tools such as SHAP facilitate clinical use by making sense of the decision-making processes of models. In this context, AI-supported approaches are ushering in a new era in disease management by improving clinical decision-making processes at all stages from pancreatic cancer diagnosis to treatment (Srinidhi & Bhargavi, 2023; Yasar, Yagin, Melekoglu, & Ardigò, 2024).

The aim of this study is to identify possible proteome biomarkers that can be used in the early diagnosis and treatment of pancreatic cancer patients with machine learning methods based on proteomic data obtained using proteomic technologies and to reveal the global effects of these biomarkers with SHap, one of the explainable artificial intelligence methods. This study is a retrospective study since the data obtained from quantitative proteomic analysis of 60 serum samples prospectively obtained from pancreatic cancer and healthy adults were used.

MATERIAL AND METHOD

Dataset

The population of this study consists of pancreatic cancer patients and healthy adults. A total of 60 serum samples (pancreatic cancer=30, control=30) were prospectively collected from these two groups. A total of 60 serum samples collected from pancreatic cancer patients and healthy adults were used in this study to identify potential biomarkers for the early diagnosis and detection of pancreatic cancer using machine learning. Serum samples were processed for enrichment of glycoproteins using Aleuria aurantia lectin (AAL) and then analyzed by both isobaric TMT proteomic labeling and label-free quantitative analysis methods. Identification and quantification of glycoproteins were performed using Orbitrap Velos and LTQ mass spectrometry, respectively. These proteomic data were then open-sourced and used in the present study (Nie et al., 2014).

Data Preprocessing and Extreme Gradient Boosting (XGBoost)

Missing values in proteomics data can occur due to low protein abundance, technical errors or below the sensitivity limit of instruments such as mass spectrometry. These missing values are particularly common in large-scale proteomic analyses and can compromise the integrity of the data. Missing value assignment is performed to prevent these omissions from affecting the accuracy and reliability of statistical analyses. A proper missing value assignment strategy is especially vital in protein-biomarker discovery, which is critical in biomedical studies (Emmanuel et al., 2021). Furthermore, given the high dimensionality of proteomics data, appropriate variable selection must be made to avoid losing important biological information; this prevents the model from overlearning and improves its overall performance, ensuring the interpretability of the results (Venkatesh & Anuradha, 2019). In this study, Random Forest was used for missing value assignment, while Elastic Net was used as a variable selection method (Jenul, Schrunner, Huynh, & Tomic, 2021; Xia et al., 2017). On the other hand, in machine learning methods, hyperparameters refer to settings outside the learning process that are set during the training of the model and directly affect the performance of the model. Hyperparameters can affect model complexity, learning speed and generalization ability. Incorrectly chosen hyperparameters can lead to overfitting or underfitting of the model. Hyperparameter optimization aims to find the optimal values of these parameters to maximize model performance. On the other hand, 70% of the data was used for training and 30% for testing the machine learning model. In this study, XGBoost (Extreme Gradient Boosting) is a

machine learning algorithm based on decision trees and was developed specifically to provide high accuracy, speed and efficiency. It is faster and memory efficient compared to other gradient boosting algorithms. XGBoost shows strong performance on large datasets and high-dimensional data. Key features of XGBoost include regularization and overfitting prevention, missing data management, parallel processing, and support for customizable objective functions (Swathi & Kodukula, 2022). Moreover, in machine learning methods, hyperparameters refer to settings outside the learning process that are adjusted during model training and directly affect the model's performance. Hyperparameters can affect model complexity, learning speed and generalization ability. Incorrectly chosen hyperparameters can lead to overfitting or underfitting of the model. Hyperparameter optimization aims to find the optimal values of these parameters to maximize model performance. In this study, a grid search algorithm with 5 repeated and 10-fold cross-validation is used for hyperparameter optimization (Shams et al., 2024). The classification performance of the model is given by AUC, F1 score, accuracy, specificity, sensitivity and brier score.

RESULTS AND DISCUSSION

Proteomic analyses of serum samples from pancreatic cancer patients and healthy adults yielded a total of 232 proteins. Among these 232 proteins, 29 proteins were selected by the Elastic Net variable selection method. Protein identification number, gene name and protein name information for 29 proteins selected by Elastic Net variable selection method are given in Table 1.

Table 1. The Information for 29 Proteins Selected By Elastic Net Variable Selection Method

Protein ID	Gene Name	Protein Name
P05543	SERPINA7	Thyroxine-binding globulin
O75636	FCN3	Ficolin-3
P25311	AZGP1	Zinc-alpha-2-glycoprotein
P02766	TTR	Transthyretin
P02750	LRG1	Leucine-rich alpha-2-glycoprotein
P14151	SELL	L-selectin
P07225	PROS1	Vitamin K-dependent protein S
P00742	F10	Coagulation factor X
P27487	DPP4	Dipeptidyl peptidase 4
P01777	IGHV3-23	Immunoglobulin heavy variable 3-23
P02748	C9	Complement component C9
P17936	IGFBP3	Insulin-like growth factor-binding protein 3
P01031	C5	Complement C5
P04275	VWF	von Willebrand factor
P01833	PIGR	Polymeric immunoglobulin receptor
P22891	PROZ	Vitamin K-dependent protein Z
P01011	SERPINA3	Alpha-1-antichymotrypsin

P06727	APOA4	Apolipoprotein A-IV
P02647	APOA1	Apolipoprotein A-I
P02649	APOE	Apolipoprotein E
P19320	VCAM1	Vascular cell adhesion protein 1
P83593	IGKV4-1	Immunoglobulin kappa variable 4-1
P05109	S100A8	Protein S100-A8
P01621	IGKV3-20	Immunoglobulin kappa variable 3-20
P02768	P02768	Albumin
P06312	IGKV4-1	Immunoglobulin kappa variable 4-1
P00739	HPR	Haptoglobin-related protein
Q15369	ELOC	Elongin-C
P06318	IGLV6-57	Immunoglobulin lambda variable 6-57

The performance metrics for the machine learning method XGBoost for classification are given in Table 2.

Table 2. The Performance Metrics for the Machine Learning Method XGBoost for Pancreatic Cancer Classification

Performance Metrics	**Value**
Area Under the Curve	0.850
Brier Score	0.154
Accuracy	0.833
Sensitivity	0.750
Specificity	0.900
F1-Score	0.800

When the performance metrics of the XGboost classification model are analyzed in Table 1, it can be said that it is quite successful in classifying pancreatic cancer based on protein data. On the other hand, the bar graph and bee swarm graph of the SHAP method applied to make the outputs of the XGBoost classification model more explainable are given in Figure 1.

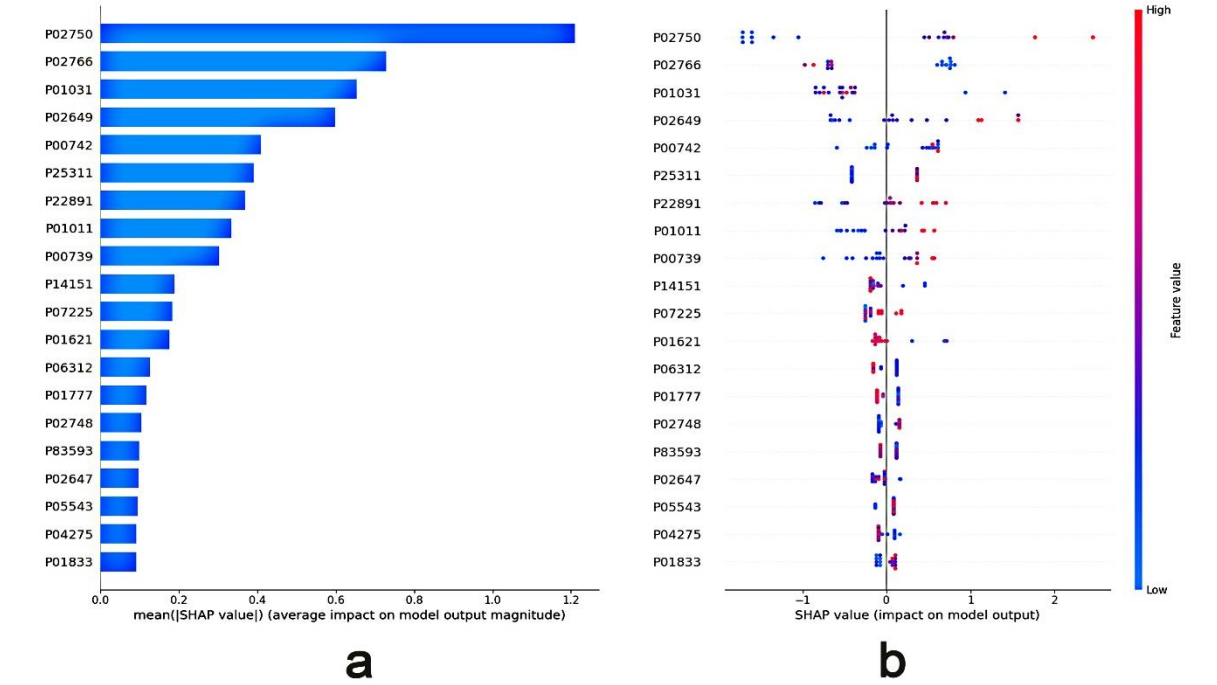

Figure 1. The Bar Graph (a) and bee Swarm Graph (b) of the SHAP Method Applied to Make the Outputs of the XGBoost Classification Model

Considering the bar graph obtained by Shap method, the four most important proteins that can be used as possible biomarkers for early diagnosis and treatment of pancreatic cancer patients are P02750 (Leucine-rich alpha-2-glycoprotein), P02766 (Transthyretin), P01031 (Complement C5) and P02649 (Apolipoprotein E). Moreover, when the bee swarm graph is analyzed, it can be said that increasing values of P02750 (Leucine-rich alpha-2-glycoprotein) and P02649 (Apolipoprotein E) proteins increase the risk of pancreatic cancer, while decreasing values of P02766 (Transthyretin) and P01031 (Complement C5) proteins increase the risk of pancreatic cancer. Similarly, increasing values of P00742 (Coagulation factor X) and P25311 (Zinc-alpha-2-glycoprotein) proteins also increase the risk of pancreatic cancer disease. Leucine-rich alpha-2-glycoprotein (LRG1) has attracted attention as a biomarker associated with pancreatic cancer. Research suggests that high serum levels of LRG1 may be associated with poor prognosis in individuals with pancreatic cancer. The effects of LRG1 on cancer cells include mechanisms such as increasing cell proliferation and inhibiting apoptosis. Furthermore, the involvement of LRG1 in inflammation and cellular response processes suggests that it may be an important factor in the development of pancreatic cancer (Lin et al., 2022; Otsuru et al., 2019). Transthyretin (TTR) is known as a protein that transports thyroid hormones and retinol (vitamin A) and has been studied in relation to pancreatic cancer. Serum TTR levels vary in individuals with pancreatic cancer, suggesting that these changes may be related to the presence and prognosis of cancer. Low TTR levels have been reported in association with pancreatic

cancer and other malignancies. Furthermore, the role of TTR in cell metabolism has led to the hypothesis that it may influence the growth and proliferation processes of pancreatic cancer cells. The association of TTR with inflammation and oxidative stress is also important, as these conditions play a critical role in the development of pancreatic cancer (Chen et al., 2013; Nanno et al., 2024). Apolipoprotein E (ApoE) is a protein that plays an important role in the metabolism of lipoproteins, and there has been some research on its association with pancreatic cancer. Studies suggest that ApoE may increase the risk or worsen the prognosis of pancreatic cancer. In particular, some genotypes of ApoE (for example, the ApoE ε4 allele) have been found to be associated with a higher risk of developing pancreatic cancer. Furthermore, the effects of ApoE on cellular signaling and inflammation may influence the growth and spread of pancreatic cancer cells. The role of ApoE in pancreatic cancer is an important area of cancer pathogenesis and therapeutic strategies (Kemp et al., 2021; Wu et al., 2024). Complement C5 is a protein that is an important component of the immune system and regulates inflammation and immune responses. Research on the relationship between pancreatic cancer and C5 suggests that this protein may play a role in cancer development and progression. In particular, the effect of C5 on inflammatory responses may affect the microenvironment of pancreatic cancer and promote the growth of tumor cells. Some studies have observed elevated levels of C5 in pancreatic cancer patients, which has been associated with a poor prognosis. In addition, the effects of C5 on various cellular signaling pathways are considered as a potential target for the development of new strategies for tumor immunotherapy (Hussain et al., 2022; Nsingwane et al., 2023).

Artificial intelligence and machine learning technologies are pioneering revolutionary changes in the healthcare sector. These technologies provide significant advances, especially in disease diagnosis and image analysis. There are many studies in this field (Develi & Sorgucu, 2015; Huang et al., 2022; Kenner et al., 2021; Qureshi, Javed, Sarmadi, Pandol, & Li, 2022; Sorgucu, Kabalcı, Develi, & Bilim, 2011). In a study aimed at increasing the understandability of decisions made by machine learning models used in sensitive medical areas such as pancreatic cancer treatment, two different models were developed, Decision Tree and Random Forest, to decide whether to administer chemotherapy to patients using 185 pancreatic cancer cases taken from The Cancer Genome Atlas Program. As a result of the study, it was seen that the Decision Tree model gave more understandable results, while the Random Forest model, although more complex, produced more accurate results. In addition, the decision-making processes of the models were made more understandable by using explainability techniques

such as SHAP and LIME. The study revealed that the patient's age, pathological M stage, tumor location and primary diagnosis type are the most important factors affecting the treatment decision and emphasized that machine learning models should be both accurate and understandable in medical decision-making processes (Bascarán et al., 2023). Another study aimed to develop an integrated method for the rapid isolation and analysis of human plasma exosomes and to distinguish different cancer types by deep learning-based MALDI-TOF MS fingerprint analysis. The researchers isolated high-purity exosomes from a single sample in a short time of 2 h using sequential size-exclusion chromatography (SSEC) technology. A total of 220 clinical samples, including 79 breast cancer patients, 57 pancreatic cancer patients, and 84 healthy controls, were analyzed in the study. After MS data preprocessing and feature selection, a multi-class artificial neural network model (Exo-ANN) was created and the model performed well in both training and test sets. An accuracy rate of 80.0% and an area under the curve (AUC) value of 0.91 were obtained for the samples in the independent test set. The results showed that the developed integrated system has the potential to be a general tool that can be used in clinical cancer diagnosis (Zheng et al., 2022). The goal of another study was to create a liquid biopsy panel with multiple biomarkers that would help with the diagnosis and staging of pancreatic ductal adenocarcinoma (PDAC) and to test how well this panel worked at making diagnoses. The research scope included the analysis of miRNA and mRNA obtained from tumor-associated outer cell vesicles, circulating cell-free DNA (cfDNA) concentration, and CA19-9 levels. We used artificial intelligence methods, particularly LASSO (Least Absolute Shrinkage and Selection Operator), to select important features in the analysis of these data. We then developed an ensemble model using multiple machine learning algorithms, including K-Nearest Neighbor, Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Logistic Regression, and Naive Bayes. We trained the model using cross-validation and bootstrap methods to prevent overfitting and enhance its generalizability. So, in a separate test group, the model was able to correctly diagnose PDAC 92% of the time, and it was also more accurate (84% of the time) than imaging methods at staging patients who did not have metastases (Yang et al., 2020).

CONCLUSION

Pancreatic cancer is one of the most lethal malignancies worldwide, with the vast majority of patients dying within the first year of diagnosis. This type of cancer presents a challenging clinical picture, as it is often diagnosed at late stages and is resistant to treatment. Current

approaches of surgery, chemotherapy and radiotherapy can only provide long-term cure in a limited group of patients and in most cases do not prevent disease progression. In recent years, research into the molecular mechanisms of pancreatic cancer has increased, with genetic and proteomic studies in particular showing promise for the discovery of new therapeutic targets. However, more studies are needed to translate these findings into clinical practice. In this context, a better understanding of pancreatic cancer biology and the development of personalized treatment approaches will play a critical role in improving patient survival rates.

In conclusion, this study successfully applied machine learning methods to identify potential proteomic biomarkers that can be used in the early diagnosis and treatment of pancreatic cancer patients. The analysis of proteomic data with XGBoost algorithm revealed that especially P02750 (Leucine-rich alpha-2-glycoprotein), P02766 (Transthyretin), P01031 (Complement C5) and P02649 (Apolipoprotein E) proteins are important biomarker candidates for pancreatic cancer diagnosis. In addition, the outputs of this model were made more understandable with the SHAP method, and how each protein affects cancer risk was explained. In particular, it was observed that increasing values of P02750 and P02649 proteins increased the risk of pancreatic cancer, while decreasing values of P02766 and P01031 proteins increased the risk. These findings may contribute to biomarker discovery and a better understanding of the molecular mechanisms of pancreatic cancer and shed light on future personalized treatment approaches. However, more large-scale confirmatory studies in different cohorts are needed for these biomarkers to be used in clinical practice.

REFERENCES

Ansari, D., Torén, W., Zhou, Q., Hu, D. & Andersson, R. (2019). Proteomic and genomic profiling of pancreatic cancer. *Cell biology and toxicology*, 35, 333-343.

Bascarán, J. B., Leal, Á. F., Mosqueira-Rey, E., Ríos, D. A., Hernández-Pereira, E. & Moret-Bonillo, V. (2023). *Understanding Machine Learning Explainability Models in the context of Pancreatic Cancer Treatment.* Paper presented at the VI Congreso XoveTIC: impulsando el talento científico.

Chen, J., Chen, L.-J., Xia, Y.-L., Zhou, H.-C., Yang, R.-B., Wu, W., . . . Zhao, Y. (2013). Identification and verification of transthyretin as a potential biomarker for pancreatic ductal adenocarcinoma. *Journal of Cancer Research and Clinical Oncology*, 139, 1117-1127.

Develi, I. & Sorgucu, U. (2015). Prediction of temperature distribution in human BEL exposed to 900 MHz mobile phone radiation using ANFIS. *Applied Soft Computing*, 37, 1029-1036.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B. & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big data*, 8, 1-37.

Halbrook, C. J., Lyssiotis, C. A., di Magliano, M. P. & Maitra, A. (2023). Pancreatic cancer: Advances and challenges. *Cell*, 186(8), 1729-1754.

Hanna-Sawires, R. G., Schiphuis, J. H., Wuhrer, M., Vasen, H. F., van Leerdam, M. E., Bonsing, B. A., . . . Tollenaar, R. A. (2021). Clinical perspective on proteomic and glycomic biomarkers for diagnosis, prognosis, and prediction of pancreatic cancer. *International journal of molecular sciences,* 22(5), 2655.

Huang, B., Huang, H., Zhang, S., Zhang, D., Shi, Q., Liu, J. & Guo, J. (2022). Artificial intelligence in pancreatic cancer. *Theranostics,* 12(16), 6931.

Hussain, N., Das, D., Pramanik, A., Pandey, M. K., Joshi, V.,& Pramanik, K. C. (2022). Targeting the complement system in pancreatic cancer drug resistance: a novel therapeutic approach. *Cancer Drug Resistance,* 5(2), 317.

Jenul, A., Schrunner, S., Huynh, B. N. & Tomic, O. (2021). RENT: A Python package for repeated elastic net feature selection. *Journal of Open Source Software,* 6(63), 3323.

Kemp, S. B., Carpenter, E. S., Steele, N. G., Donahue, K. L., Nwosu, Z. C., Pacheco, A., . . . The, S. (2021). Apolipoprotein E promotes immune suppression in pancreatic cancer through NF-κB–mediated production of CXCL1. *Cancer research,* 81(16), 4305-4318.

Kenner, B., Chari, S. T., Kelsen, D., Klimstra, D. S., Pandol, S. J., Rosenthal, M., . . . Abul-Husn, N. (2021). Artificial intelligence and early detection of pancreatic cancer: 2020 summative review. *Pancreas,* 50(3), 251-279.

Kolbeinsson, H. M., Chandana, S., Wright, G. P. & Chung, M. (2023). Pancreatic cancer: a review of current treatment and novel therapies. *Journal of Investigative Surgery,* 36(1), 2129884.

Lin, M., Liu, J., Zhang, F., Qi, G., Tao, S., Fan, W., . . . Zhou, F. (2022). The role of leucine-rich alpha-2-glycoprotein-1 in proliferation, migration, and invasion of tumors. *Journal of Cancer Research and Clinical Oncology*, 1-9.

Mizrahi, J. D., Surana, R., Valle, J. W. & Shroff, R. T. (2020). Pancreatic cancer. *The Lancet,* 395(10242), 2008-2020.

Nanno, Y., Toyama, H., Mizumoto, T., Ishida, J., Urade, T., Fukushima, K., . . . Asari, S. (2024). Preoperative level of serum transthyretin as a novel biomarker predicting survival in resected pancreatic ductal adenocarcinoma with neoadjuvant therapy. *Pancreatology,* 24(6), 917-924.

Nie, S., Lo, A., Wu, J., Zhu, J., Tan, Z., Simeone, D. M., . . . Lubman, D. M. (2014). Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. *Journal of Proteome Research,* 13(4), 1873-1884.

Nsingwane, Z., Naicker, P., Omoshoro-Jones, J., Devar, J., Smith, M., Candy, G., . . . Nweke, E. E. (2023). Inhibition of the Complement Pathway Induces Cellular Proliferation and Migration in Pancreatic Ductal Adenocarcinoma. *medRxiv*, 2023.2008. 2008.23293417.

Otsuru, T., Kobayashi, S., Wada, H., Takahashi, T., Gotoh, K., Iwagami, Y., . . . Serada, S. (2019). Epithelial-mesenchymal transition via transforming growth factor beta in pancreatic cancer is potentiated by the inflammatory glycoprotein leucine-rich alpha-2 glycoprotein. *Cancer science,* 110(3), 985-996.

Qureshi, T. A., Javed, S., Sarmadi, T., Pandol, S. J. & Li, D. (2022). Artificial intelligence and imaging for risk prediction of pancreatic cancer. *Chinese clinical oncology,* 11(1), 1.

Shams, M. Y., Elshewey, A. M., El-Kenawy, E.-S. M., Ibrahim, A., Talaat, F. M. & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications,* 83(12), 35307-35334.

Sorgucu, U., Kabalcı, Y., Develi, İ. ve Bilim, M. (2011). *Aşırı Düşük Frekanslı Elektromanyetik Alanın Derideki Hidroksiprolin Seviyesi Üzerine Olan Etkisinin ANFIS Kullanılarak Modellenmesi.* Paper presented at the 6th International Advanced Technologies Symposium (IATS'11).

Srinidhi, B. & Bhargavi, M. (2023). *An XAI Approach to Predictive Analytics of Pancreatic Cancer.* Paper presented at the 2023 International Conference on Information Technology (ICIT).

Swathi, K. & Kodukula, S. (2022). XGBoost Classifier with Hyperband Optimization for Cancer Prediction Based on Geneselection by Using Machine Learning Techniques. *Revue d'Intelligence Artificielle, 36*(5).

Venkatesh, B. & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and information technologies,* 19(1), 3-26.

Wu, B., Yang, X., Chen, F., Song, Z., Ding, X. & Wang, X. (2024). Apolipoprotein E is a prognostic factor for pancreatic cancer and associates with immune infiltration. *Cytokine,* 179, 156628.

Xia, J., Zhang, S., Cai, G., Li, L., Pan, Q., Yan, J. & Ning, G. (2017). Adjusted weight voting algorithm for random forests in handling missing values. *Pattern Recognition,* 69, 52-60.

Yang, Z., LaRiviere, M. J., Ko, J., Till, J. E., Christensen, T., Yee, S. S., . . . Shen, H. (2020). A multianalyte panel consisting of extracellular vesicle miRNAs and mRNAs, cfDNA, and CA19-9 shows utility for diagnosis and staging of pancreatic ductal adenocarcinoma. *Clinical Cancer Research,* 26(13), 3248-3258.

Yasar, S., Yagin, F. H., Melekoglu, R. & Ardigò, L. P. (2024). Integrating proteomics and explainable artificial intelligence: a comprehensive analysis of protein biomarkers for endometrial cancer diagnosis and prognosis. *Frontiers in Molecular Biosciences,* 11, 1389325.

Zheng, H., Zhao, J., Wang, X., Yan, S., Chu, H., Gao, M. & Zhang, X. (2022). Integrated pipeline of rapid isolation and analysis of human plasma exosomes for cancer discrimination based on deep learning of MALDI-TOF MS fingerprints. *Analytical Chemistry,* 94(3), 1831-1839.