Stacking Ensemble Learning of Extracellular Vesicle Dynamics Reveals Disease Aggressiveness in Pancreatic Cancer

Angeliki Iosif, Vasileios E. Katsigiannis, Ioanna Angelioudaki, Vassiliki Kigka, Alexandros Georgios Tzingounis, Stavros T. Miloulis, Myrto Costopoulos, Androniki-Maria Skreka, Nikolaos Memos, Agapi Kataki, George K. Matsopoulos, Manousos M. Konstadoulakis, Dimitrios I. Fotiadis, *Fellow IEEE*, Konstadina Kourou

***Abstract*— Pancreatic ductal adenocarcinoma (PDAC) is an aggressive malignancy with extremely poor prognosis. Although tissue biopsy before surgery is a common prerequisite and contributes to an initial diagnosis, it does not provide any information about the aggressiveness of the tumor, which is critical for patient management. These data can only be found on the histology report obtained later after surgery. As liquid biopsy could provide informative molecular data much faster, it represents an appropriate candidate to bridge this gap. However, the complexity of data requires the use of advanced machine learning (ML) models for accurate prediction. The present study employs a comprehensive data-driven pipeline to address this important issue. The ML-based pipeline entailed the appropriate steps for unbiased model training and testing. EV size and surface protein markers quantified via flow cytometry, as CD45, in combination with clinical and laboratory data served as input data. According to our findings, $CD45^-$ EV populations outperformed $CD45^+$(0.63) and the entire dataset with an AUC of 0.73, indicating that tumor-derived EV subpopulations are more informative than immune-derived ones. These findings highlight the feasibility of incorporating adaptive learning analysis and prediction models into routine clinical procedures and demonstrate the potential of EV-derived features.**

***Clinical Relevance*— The integration of EV-derived biomarkers with ensemble learning techniques may enable minimally invasive assessment of PDAC aggressiveness, facilitating earlier risk stratification and personalized treatment strategies both before and after surgery.**

Keywords—Pancreatic ductal adenocarcinoma, tumor aggressiveness, extracellular vesicles, machine learning, photonic data measurements

I. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most prevalent subtype of pancreatic cancer and remains one of the most lethal human malignancies, with a five-year overall survival rate of approximately 10%. Its poor prognosis is largely driven by a dense desmoplastic stroma and an immunosuppressive tumor microenvironment that hinder drug delivery and promote progression [1]. Histological confirmation obtained through tissue biopsy represents the current gold standard for PDAC diagnosis. However, biopsy-based assessment of PDAC has notable limitations, as it provides only partial information on tumor aggressiveness, including perivascular/lymphovascular invasion and vascular infiltration [2], [3]. These critical pathological parameters are typically available only postoperatively through comprehensive histopathological evaluation, delaying risk stratification, and therapeutic decision-making.

Liquid biopsy has emerged as a promising minimally invasive alternative to address these limitations. Among liquid biopsy modalities, extracellular vesicles (EVs) have gained considerable attention due to their abundance in biological fluids, stability in circulation, and capacity to carry tumor-derived nucleic acids and proteins that reflect the molecular and biological characteristics of PDAC. EVs comprise a heterogeneous population of membrane-bound particles, ranging from small EVs (<500nm) to very large EVs (up to 10μm) [4], which are increasingly recognized as key mediators of intercellular communication within the tumor microenvironment and at distant sites. In this study, we focus on large EVs in the 2-5μm range, which are tractable by standard flow cytometry. Machine learning (ML) approaches have been widely applied across various aspects of cancer management, owing to their ability to learn directly from complex data and to support personalized treatment strategies [5]. In PDAC, existing studies have primary investigated

* Research funding from the European Union's HORIZON-CL4-2022-DIGITAL-EMERGING-01 project under GA nos.1010931712020 Institutional Review Board Statement (https://horizon-de-phorever.eu/).

K. K., A. I., V. K. and D. I. F are with the Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR-45110 GREECE and with the Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, GR-45110, GREECE (e-mail: k.kourou@uoi.gr, e-mail: agg.iosif@uoi.gr, email: kigkavaso@gmail.com , e-mail: fotiadis@uoi.gr , tel: +30 265100 7298).

I. A., A. G. T., A. M. S., and M. M. K. are with the 2nd Department of Surgery, Aretaieion Hospital, Medical School of Athens, National and Kapodistrian University of Athens, Athens, Greece (e-mail: ioangel95@gmail.com, email: alextzingounis@hotmail.com, email: nikiskre@gmail.com, e-mail: konstadoulakismm@yahoo.com).

A. K. and N. M. are with the 1st Department of Propaedeutic Surgery, "Hippokratio" General Hospital, Medical School of Athens National and Kapodistrian University of Athens, Greece (e-mail: akataki@med.uoa.gr, email: nikolaosmemos@gmail.com).

M.C is with Flow Cytometry Department, BioAnalytica SA, Athens, Greece (email: m.kostopoulou@bioanalytica.gr).

V. E. K., S. T. M., and G. M. are with Biomedical Engineering Laboratory School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (email: vassiliskat@biomed.ntua.gr, email: smiloulis@biomed.ntua.gr, email: gmatsopoulos@biomed.ntua.gr).

tumor aggressiveness through survival analysis or radiomic feature analysis. Furthermore, no study to date has combined photonics-based large EV quantification with ensemble learning to predict perivascular infiltration in PDAC, which remains a critical aggressiveness marker only available post-surgery. The current work is one of the first to use adaptive learning algorithms to assess PDAC aggressiveness utilizing plasma-derived large EVs (2 to 5μm) as a cost-effective and rapid screening method. This methodology combines EV-based characteristics with clinical data using an adaptable ensemble learning framework. The model's performance was rigorously evaluated to assess its ability to identify aggressive tumor features such as perivascular infiltration, allowing for deeper insights into the disease's biology.

II. Materials and Methods

A. Study population

Ninety-eight patients with PDAC were enrolled, treated in two Surgical Clinics of National and Kapodistrian University of Athens (NKUA) from July 2020 to December 2024. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics and Deontology Committees (237/10-7-2020, 34/14-7-2020, 488/20-02-2023). All patients provided written informed consent for the assessment of their data for research purposes. Inclusion criteria involved histopathologically confirmed PDAC diagnosis. Clinical and laboratory features were retrieved from patients' medical files.

B. Flow Cytometry (FCM)

All patients followed a fasting diet before blood drawing. Regarding photonic data extraction, plasma samples were subjected to sequential centrifugation to remove cells and cellular debris, after which aliquots were processed using a BD FACSLyric™ Flow Cytometry System. Intact EVs were evaluated using a custom kit (BD 626267) in combination with Trucount™ Tubes (BD Biosciences) for quantification, following the manufacturer's instructions. EV sizing was performed using the Rosetta Calibration Beads mix and software (Exometry Inc., Amsterdam, Netherlands). Analyses focused on intact EVs within three approximate size ranges: 2μm, 3μm, and 5μm. EVs were stained with the following antibodies: hematopoietic marker CD45 (PerCP-Cy5.5 Mouse Anti-Human CD45; BD 564105), tetraspanin CD63 commonly found on EVs (PE Mouse Anti-Human CD63; BD 556020), and receptor tyrosine kinase EphA2 frequently overexpressed in PDAC (BV421 Mouse Anti-Human Epha2; BD OptiBuild 748144). Flow cytometry thresholds were set according to the manufacturer's recommendations, and samples were acquired at medium flow rate.

C. Predictor and outcome variables

Analysis was based on demographic data and clinical features obtained from PDAC patients right before the surgery. EV-based characteristics refer to the expression of CD45, CD63, and EPHA2 antibodies. Tumor aggressiveness was assessed based on perivascular infiltration as the primary outcome. Absence of perivascular infiltration was annotated as *Class 0* (n=52) while its presence as *Class 1* (n=46).

D. Data analysis

In the current study, a stacking ensemble approach [6] was used to address disease aggressiveness and identify relevant predictors. The analysis pipeline entailed preprocessing steps, feature selection, model training, and testing and hyperparameter tuning. A nested cross-validation (5-fold CV) strategy was applied to the ensemble methodology to separate hyperparameter tuning, in the inner loop, from generalization error estimation, in the outer loop, avoiding the use of the same data for both steps. This design is particularly critical given the relatively small sample size of the dataset, as standard k-fold CV has been shown to produce substantially biased performance estimates in such settings, while nested CV provides an almost unbiased estimate of true error [7] [8]. Feature selection was incorporated within the inner loop to prevent any information leakage from test data. The scikit-learn library for ML in Python [9] was used for building the proposed integrated ML-based pipeline.

To choose only the important features for training and testing the best performing model, a meta-estimator with the Random Forest (RF) algorithm [9] was integrated within the pipeline, with reference to the selected base estimators. The performance was assessed on the test set using the area under the Receiver Operating Characteristic curve (ROC AUC), which involved assessing additional metrics: accuracy, specificity, sensitivity, and AUC. Raw data were initially rescaled to zero mean and unit variance. Missingness was below 30% across all cases and variables; the remaining missing values imputed using k-nearest neighbors (KNN) imputer. KNN imputation was selected for its ability to handle mixed-type variables and leverage local data structure.

E. Feature Selection

A meta-transformer built on a Random Forest (RF) algorithm was used for feature selection [9] [10]. This meta-transformer assigns weights to the features and ranks them according to their relative importance. The feature selection step was incorporated into the ML-based pipeline alongside the stacking model to select the relevant features for training the model. The maximum number of features selected by the estimator was restricted to 10 to identify the most informative predictors that contribute to PDAC aggressiveness.

F. Model training and validation

Nested CV was applied to train a model with optimized parameters. In the inner loop the score was approximately maximized by fitting a model to each training set. In the outer loop generalization error was estimated by averaging test set scores over several dataset splits. Hence, both generalization errors and misclassifications during the training phase were reduced. Class imbalance handling was addressed using undersampling technique (Random UnderSampler via Imbalanced-learn) [11]. This approach downscales the majority class randomly picking samples with or without replacement. Class imbalance handling was considered only during the training phase to avoid any bias during the testing of the prediction model. Undersampling was preferred to maintain the integrity of the original dataset. Although the feature types in the current dataset is mixed, with continuous and categorical variables, the vast majority are numerical.

The stacking ensemble framework incorporated heterogeneous base learners, including distance-based, probabilistic, and tree-based classifiers. Stacking employs a two-level architecture in which the first-level base models are trained on the original training data, and a second-level-meta-model is subsequently trained on the predictions by these base models. The base models are: (i) KNeighbors Classifier, (ii) Gaussian Naïve Bayes, (iii) Random Forest Classifier, and (iv) XGBoost Classifier [6]. Logistic regression was chosen as a meta-model because of its simplicity, interpretability, and low tendency to overfit in small datasets. Other models (such as Balanced Random Forest) were also tested, while stacking yielded improved performance. Hyperparameter tuning was applied to every estimator using the grid search procedure to identify the optimal model parameters through exhaustive evaluation of predefined parameter combinations. The following metrics were calculated to assess the performance of the classification model: specificity (true negative rate); sensitivity (true positive rate); AUC, F1score, and accuracy. The ROC curve was also calculated to depict the trade-off between false negative and false positive rates at each possible cut-off. Explainability analysis was performed using SHAP values to identify the most significant features that drive the models' predictions [12]. A positive SHAP number indicates that the specific trait improves the prediction, whilst a negative value suggests the opposite. This analysis enhances understanding of the role each feature plays in shaping the model's behavior for a given prediction.

III. Results

Analysis was conducted on three distinct datasets and use cases: (i) the complete dataset incorporating all EV populations; (ii) a dataset restricted to $CD45^+$ EV populations; and (iii) a dataset including only $CD45^-$ EV populations. Clinical and laboratory features were included in all cases. Table I presents the evaluation metrics for each use case. Sections A and B below present the results for the two use cases based on separate models trained on $CD45^+$ EVs and $CD45^-$ EVs respectively. Interestingly, the total dataset model, which combines both $CD45^+$ and $CD45^-$ EV subpopulations, performed identically to the $CD45^+$ model alone (AUC = 0.63), indicating that the inclusion of $CD45^+$ EV features dilutes rather than augments the discriminative signal carried by the $CD45^-$ subpopulation.

TABLE I. PERFORMANCE OF THE STACKED MODEL IN PREDICTING PERIVASCULAR INFILTRATION

Use cases	Evaluation Metrics				
	Accuracy	Specificity	Sensitivity	F1score	AUC
Total dataset	0.63	0.62	0.64	0.62	0.63
$CD45^+$	0.63	0.62	0.64	0.62	0.63
$CD45^-$	0.73	0.68	0.78	0.73	0.73

A. *$CD45^+$ EVs populations*

The stacked model demonstrated similar performance of $CD45^+$ EV populations and the complete dataset (Table I). Despite the moderate accuracy score of 0.63 in the $CD45^+$ model, it is important to examine the features underlying this performance and to determine whether EV-related variables rank among the most influential predictors. Eight out of ten highest-ranked features were derived from EV-based measurements, while the remaining two (amylase and Carcinoembryonic Antigen (CEA)) originated from clinical and laboratory variables (Fig. 1). The EV-related features corresponded to vesicle populations of all tested size ranges. Among the top ranked features is the combined expression of $CD45^+$, $CD63^+$, and $EPHA2^+$ markers, demonstrating a clinically meaningful association with the predicted outcome.

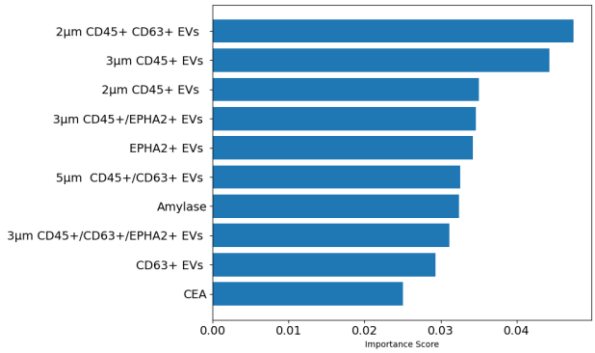

Fig. 1. Bar plot of the top 10 features for the stacked model of the first use case ranked by relative importance. Features are ranked on the y-axis by descending order of importance, with corresponding scores on the x-axis.

B. *$CD45^-$ EVs populations*

This use case outperformed the others with an accuracy score equal to 0.73. EV-derived features (7/10) dominated the ranking, especially EV populations expressing $CD45^-/EPHA2^+$, and $CD45^-/CD63^+$ markers across all EVs size ranges (Fig. 2). High values of $CD45^-/EPHA2^+$ EVs, especially at 2µm and 3µm, were associated with positive SHAP values, indicating a strong contribution toward increasing the model's predicted outcome. Clinical and laboratory variables, including CEA, Cancer Antigen 19.9 (CA19.9), and Mean Corpuscular Volume (MCV), also contributed to the model, though with comparatively smaller effect sizes.

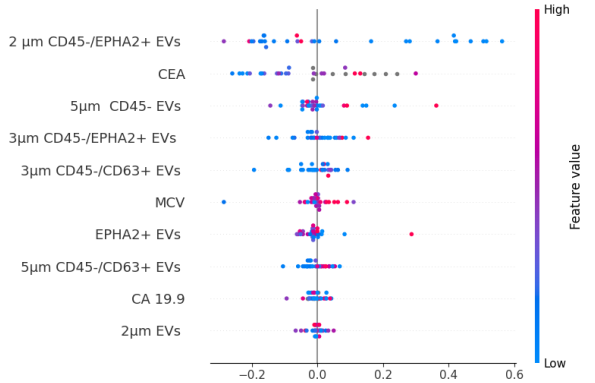

Fig.2. SHAP summary plot illustrates the contribution of the top ten features to the stacked model's prediction for the second use case. Points denote individual samples, SHAP values are shown on the x-axis, and features are ranked by mean absolute SHAP value, with color indicating low (blue) to high (red) feature values.

IV. Discussion

The present study used an ensemble stack model to predict disease aggressiveness through perivascular infiltration in PDAC patients based on EV features and relevant clinical variables. As a target of tumor aggressiveness, perivascular infiltration is characterized as a highly coordinated process

that is driven by tumor-endothelial crosstalk, extracellular matrix remodeling, and directional migration along vascular structures [13].

This finding is consistent with the established role of CD45 as a marker of hematopoietic origin in EV biology. $CD45^-$ EVs in plasma are by definition depleted of leukocyte-derived particles and thus more likely to reflect non-hematopoietic, tumor-derived components of the tumor microenvironment [14], whereas $CD45^+$ EVs predominantly capture systemic immune activity that, while prognostically relevant in its own right, appears to introduce heterogeneity that is orthogonal to the local vascular invasion dynamics underlying perivascular infiltration. The total dataset model performed no better than the $CD45^+$ model alone, and significantly worse than the $CD45^-$ model. This supports the interpretation that immune-derived EV signatures introduce noise into the prediction of perivascular infiltration, effectively masking the tumor-derived signal. This highlights the importance of EV subpopulation stratification as a preprocessing step when applying machine learning to liquid biopsy data in PDAC.

EV signature has been previously used as a prognostic factor for non-invasive risk stratification and survival prediction of PDAC patients [15]. Additionally several studies have explored its use for PDAC diagnosis using adaptive learning analysis [16] [14]. Furthermore, EV-based attempts have been implemented to predict aggressiveness focused on perineural invasion [17]. The current study extends prior research on tumor aggressiveness by concentrating on perivascular infiltration. The generalizability of the prediction model was assessed in cases that were not considered during the training phase, to avoid overfitting and reduce classification errors. The ensemble model achieved a better accuracy score when applied to the $CD45^-$ EV subpopulation compared with the full dataset.

Further investigation of these $CD45^-$ EVs populations may improve understanding of PDAC aggressiveness and metastasis while supporting their potential use in risk stratification and treatment decisions before and after surgery. However, these findings are limited by the small sample size and the absence of longitudinal follow-up data. To address challenges such as missing values in the datasets, future work may focus on the integration of synthetic data generation techniques to augment incomplete cases and improve model robustness, while evaluating the similarity and fairness of synthetic and real data to prevent bias into the analysis [18]. The inclusion of fairness assessments into data preprocessing and model evaluation, while also utilizing bias mitigation across subgroups, will help ensure non-biased and generalizable results.

V. Conclusions

This study demonstrates that photonics-derived big EV characteristics, specifically $CD45^-$ subpopulations, can predict perivascular infiltration in PDAC with an AUC of 0.73 utilizing a stacking ensemble framework used to a minimally invasive liquid biopsy. $CD45^-$ EVs are consistently identified as the most informative predictor across models, supporting their biological relevance as circulating reporters of local tumor invasiveness. This establishes EV subpopulation stratification as a meaningful preprocessing decision in ML-based liquid biopsy pipelines. These findings provide the groundwork for incorporating EV-derived ML models into clinical workflows, providing oncologists with fast, objective tool for pre- and postoperative patient risk classification, as well as tailoring personalized treatment options in PDAC.

References

[1] Y. Masugi, "The Desmoplastic Stroma of Pancreatic Cancer: Multilayered Levels of Heterogeneity, Clinical Significance, and Therapeutic Opportunities," *Cancers*, vol. 14, no. 13, Jul. 2022, doi: 10.3390/cancers14133293.

[2] S. Zhang, D. Li, Y. Liu, C. Qin, L. Tong, and L. Xu, "Multifunctional exosome-driven pancreatic cancer diagnostics and therapeutics," *Extracellular Vesicle*, vol. 2, p. 100022, Dec. 2023, doi: 10.1016/j.vesic.2023.100022.

[3] Z. Lin *et al.*, "Early Detection of Pancreatic Cancer: Current Advances and Future Opportunities," *Biomedicines*, vol. 13, no. 7, Jul. 2025, doi: 10.3390/biomedicines13071733.

[4] M. A. Khan, S. Anand, S. K. Deshmukh, S. Singh, and A. P. Singh, "Determining the Size Distribution and Integrity of Extracellular Vesicles by Dynamic Light Scattering," *Methods Mol Biol*, vol. 2413, pp. 165–175, 2022, doi: 10.1007/978-1-0716-1896-7_17.

[5] M. Wang, W. Chang, and Y. Zhang, "Artificial Intelligence for the Diagnosis and Management of Cancers: Potentials and Challenges," *MedComm (2020)*, vol. 6, no. 11, p. e70460, Nov. 2025, doi: 10.1002/mco2.70460.

[6] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[7] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLOS ONE*, vol. 14, no. 11, p. e0224365, Nov. 2019, doi: 10.1371/journal.pone.0224365.

[8] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, p. 91, Feb. 2006, doi: 10.1186/1471-2105-7-91.

[9] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 2825–2830, Nov. 2011.

[10] S. Sumathi and S. N. Sivanandam, *Introduction to Data Mining and its Applications*, vol. 29. in Studies in Computational Intelligence, vol. 29. Berlin, Heidelberg: Springer, 2006. doi: 10.1007/978-3-540-34351-6.

[11] T. Wongvorachan, S. He, O. Bulut, T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 1, Jan. 2023, doi: 10.3390/info14010054.

[12] S. M. Lundberg *et al.*, "From Local Explanations to Global Understanding with Explainable AI for Trees," *Nat Mach Intell*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[13] A. Nowosad, J.-C. Marine, and P. Karras, "Perivascular niches: critical hubs in cancer evolution," *Trends Cancer*, vol. 9, no. 11, pp. 897–910, Nov. 2023, doi: 10.1016/j.trecan.2023.06.010.

[14] I. Angelioudaki *et al.*, "A machine-learning approach for pancreatic neoplasia classification based on plasma extracellular vesicles," *Front. Oncol.*, vol. 15, Apr. 2025, doi: 10.3389/fonc.2025.1540195.

[15] Y. Han *et al.*, "Plasma extracellular vesicle messenger RNA profiling identifies prognostic EV signature for non-invasive risk stratification for survival prediction of patients with pancreatic ductal adenocarcinoma," *J Hematol Oncol*, vol. 16, no. 1, p. 7, Feb. 2023, doi: 10.1186/s13045-023-01404-w.

[16] Z. Liu, S. Jia, and L. Cao, "Using Machine Learning Methods to Develop Diagnostic and Prognostic mRNA Signatures for Pancreatic Cancer in Plasma Small Extracellular Vesicles," *Dig Dis Sci*, vol. 70, no. 7, pp. 2381–2394, Jul. 2025, doi: 10.1007/s10620-025-08867-6.

[17] Y. Sun, Y. Li, M. Li, T. Hu, and J. Wang, "Radiomics analysis using machine learning to predict perineural invasion in pancreatic cancer," *BMC Cancer*, vol. 25, p. 1480, Sep. 2025, doi: 10.1186/s12885-025-14806-5.

[18] R. J. Chen *et al.*, "Algorithm fairness in artificial intelligence for medicine and healthcare," *Nature biomedical engineering*, vol. 7, no. 6, p. 719, Jun. 2023, doi: 10.1038/s41551-023-01056-8.