RESEARCH ARTICLE

ADVANCED
INTELLIGENT
SYSTEMS
Open Access

www.advintellsyst.com

Development of a Machine-Learning Model for Diagnosis of Pancreatic Cancer from Serum Samples Analyzed by Thermal Liquid Biopsy

Sonia Hermoso-Durán, Nicolas Fraunhoffer, Judith Millastre-Bocos, Oscar Sanchez-Gracia, Pablo F. Garrido, Sonia Vega, Ángel Lanas, Juan Iovanna, Adrián Velázquez-Campoy,* and Olga Abian*

Pancreatic ductal adenocarcinoma (PDAC) poses a considerable diagnostic and therapeutic challenge due to the lack of specific biomarkers and late diagnosis. Early detection is crucial for improving prognosis, but current techniques are insufficient. An innovative approach based on differential scanning calorimetry (DSC) of blood serum samples, thermal liquid biopsy (TLB), combined with machine-learning (ML) analysis, may offer a more efficient method for diagnosing PDAC. Serum samples from a cohort of 212 PDAC patients and 184 healthy controls are studied. DSC thermograms are analyzed using ML models. The generated models are built applying algorithms based on penalized regression, resampling, categorization, cross validation, and variable selection. The ML-based model demonstrates outstanding ability to discriminate between PDAC patients and control subjects, with a sensitivity of 90% and an area under the ROC receiver operating characteristic curve of 0.83 in the training and test groups. Application of the model to an independent validation cohort of 113 PDAC patients confirms its robustness and utility as a diagnosis tool. The application of ML to serum TLB data emerges as a promising methodology for early diagnosis, representing a significant advance for detecting and managing PDAC, envisaging a minimally invasive and more efficient methodology for identifying biomarkers.

1. Introduction

Pancreatic cancer stands as a major global health issue, characterized by its high mortality rate and escalating incidence. The American Cancer Society's estimates about 64 050 new cases and 50 550 deaths in its 2023 report for US, making up ≈3% of all cancers and 7% of cancer deaths. This trend reflects the global patterns reported by the World Health Organization (WHO) and Global Cancer Observatory (GLOBOCAN), which indicate a steady rise in pancreatic cancer incidence, linked to aging populations, lifestyle changes, and risk factors such as smoking and obesity.[1,2]

Notably, regional variations in pancreatic cancer incidence and mortality suggest the influence of genetic, dietary, and environmental factors. Developed nations often report higher cancer rates, potentially due to greater exposure to risk factors and more robust cancer registries. In Europe,

S. Hermoso-Durán, P. F. Garrido, S. Vega, A. Velázquez-Campoy, O. Abian
Institute of Biocomputation and Physics of Complex Systems (BIFI)
Joint Units IQFR-CSIC-BIFI, and GBsC-CSIC-BIFI
Universidad de Zaragoza
50018 Zaragoza, Spain
E-mail: adrianvc@unizar.es; oabifra@unizar.es

S. Hermoso-Durán, Á. Lanas, A. Velázquez-Campoy, O. Abian
Instituto de Investigación Sanitaria Aragón (IIS Aragón)
50009 Zaragoza, Spain

S. Hermoso-Durán, A. Velázquez-Campoy, O. Abian
Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd)
28029 Madrid, Spain

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202400308.

N. Fraunhoffer, J. Iovanna
Aix-Marseille Université, CNRS, Centre Interdisciplinaire de Nanoscience de Marseille, Equipe Labélisée Ligue Nationale Contre le Cancer
13288 Marseille, France

J. Millastre-Bocos, Á. Lanas
Hospital Clínico Universitario Lozano Blesa
50009 Zaragoza, Spain

O. Sanchez-Gracia
Departamento de Ingeniería Electrónica y Comunicaciones
Universidad de Zaragoza
50018 Zaragoza, Spain

P. F. Garrido
Department of Physics
University of Oslo
0316 Oslo, Norway

P. F. Garrido
Center for Lifespan Changes in Brain and Cognition
University of Oslo
0371 Oslo, Norway

variations are evident across the continent, with Eastern Europe showing higher rates, possibly influenced by lifestyle and healthcare disparities.[3]

These patterns underscore the urgency of targeted prevention strategies and research to better understand the disease's pathophysiology. Implementing effective screening programs for at-risk populations could improve early detection rates and, consequently, treatment outcomes for pancreatic cancer.[1]

One of the primary challenges in diagnosing pancreatic cancer is the absence of specific symptoms in its early stages, often rendering it undetectable until it has advanced too far.[1,4] The manifestation of symptoms largely depends on the location, size, and spread of the tumor. Common areas affected by metastasis include the liver, peritoneum, lungs, and bones. This late diagnosis, due to nonspecific or absent early symptoms, significantly impacts treatment options and prognosis, highlighting the urgent need for improved diagnostic methods.[3]

While there are no general pancreatic cancer detection programs that have proven useful in the general population, imaging tests are mainly applied in high-risk groups, with low-quality evidence in relation to their efficiency, with current biomarkers being not useful as early diagnosis. For this reason, research is increasingly focusing on high-risk groups.[1] These include individuals with a family history of pancreatic cancer or genetic syndromes like Peutz–Jeghers syndrome and hereditary pancreatitis. For these groups, the use of regular imaging tests, such as endoscopic ultrasound (EUS) or magnetic resonance imaging (MRI), combined with biomarkers, is being explored to improve early detection.[1] Furthermore, advancements in genomic sequencing are enhancing our ability to identify individuals at increased risk, allowing for more targeted screening approaches. This shift toward personalized screening reflects a nuanced understanding of the disease's risk factors and underscores the need for ongoing research in biomarker development.[5]

In the study of pancreatic ductal adenocarcinoma (PDAC), the most prevalent type of pancreatic neoplasm, standard blood tests including biochemistry markers, complete blood count, and coagulation profile play a crucial role. The principal biomarker is the carbohydrate antigen 19-9 (CA19.9). This tumor marker is not exclusive to pancreatic cancer, but it is used for prognostic and follow-up purposes. Its utility is limited in asymptomatic individuals due to low negative predictive value. Genetic factors, such as the Lewis negative genotype, can affect CA19.9 production, influencing its effectiveness as a diagnostic tool.[6–8] These biomarkers, while useful, have limitations, and, therefore, there is a need for more specific and sensitive markers for PDAC.

J. Iovanna
Hospital de Alta Complejidad El Cruce
Florencio Varela, Buenos Aires B1888, Argentina

J. Iovanna
University Arturo Jauretche
Florencio Varela, Buenos Aires B1888, Argentina

A. Velázquez-Campoy, O. Abian
Departamento de Bioquímica y Biología Molecular
Universidad de Zaragoza
50009 Zaragoza, Spain

Research must continue by exploring novel biomarkers toward more effective diagnosis and management of pancreatic cancer.

Imaging techniques play a pivotal role in diagnosing PDAC, enabling assessment of tumor extent and treatment planning. Commonly used modalities include abdominal ultrasound, computed tomography, MRI, and EUS. These techniques facilitate PDAC staging using the system developed by the American Joint Committee on Cancer (AJCC),[9] considering tumor size (T), nodal involvement (N), and distant metastasis (M). Imaging reveals tumor size and metastasis, while nodal spread is typically confirmed surgically by anatomopathological analysis of the surgical specimen. The eighth edition of AJCC (2017) outlines stages IA to IV based on these criteria, aiding in therapeutic decision-making and prognostication. This imaging-based classification is crucial for tailoring treatment approaches to each cancer stage.

Beyond the stage classification, PDAC is also classified from a surgical perspective into resectable and unresectable categories. Resectable PDAC is confined to the pancreas or has limited spread and can be entirely removed. Unresectable PDAC, subdivided into locally advanced and metastatic, cannot be fully removed by surgery. Locally advanced PDAC has grown around major blood vessels but has not spread to distant organs. Metastatic PDAC involves distant organ spread.

However, challenges remain in early detection and precise monitoring due to the lack of highly sensitive biomarkers, emphasizing the need for developing noninvasive or minimally invasive techniques, such as blood sample analysis for improved diagnosis and patient management.

Peripheral blood samples represent another valuable tool in clinical evaluation for diagnosing and monitoring diseases, including PDAC. They influence about 60%–70% of medical decisions due to their ease of collection and minimal invasiveness. Blood reflects nearly everything happening in the body, containing more than 3000 proteins and peptides. Despite the dominance of a few proteins in the blood proteome, the remaining 1% includes low-abundance proteins with strong disease correlation. Advances in mass spectrometry and electrophoresis have highlighted the relevance of minor components, known as the peptidome. The interactome theory suggests that these potential biomarkers form complexes with abundant proteins, complicating their study but highlighting the multi-protein involvement in disease states. This complexity necessitates innovative biochemical and analytical approaches for successful biomarker discovery, as no single method suffices using current technologies. Differential scanning calorimetry (DSC) offers a novel analytical biophysical approach to study the proteome and peptidome based on sample thermal stability.[10]

DSC is a biophysical technique traditionally used to study protein thermal stability, which has been proven invaluable in providing information on many diseases. In DSC, protein samples are heated in a controlled environment, allowing observation of unfolding processes. The DSC thermogram of a single protein informs about its structural stability which, for certain proteins, can be altered by the effect of a disease. This method provides unique insights into the protein composition of biological samples, such as blood serum or plasma, revealing potential disease markers. Changes in the DSC thermogram of blood serum/plasma may indicate alterations in protein composition and/or

interactions, which are significant in understanding disease pathology and progression. While not directly identifying low-abundance biomarkers, DSC offers relevant information about the overall state of the proteome, making it a promising tool for disease diagnosis and monitoring.[11]

Recently, thermogram analysis in DSC has evolved significantly, particularly in diagnosing and monitoring of tumoral diseases.[12–17] This technique, named as thermal liquid biopsy (TLB), involves comparing thermograms from healthy individuals and patients to extract specific, differential thermodynamic parameters.[10,16,18–32] Notable developments include nonparametric methods for thermogram classification, high-dimensional data analysis, and algorithms that adjust for variable penalization.[33–38] These approaches have been applied to a range of diseases, demonstrating the versatility of TLB in clinical diagnostics. The ongoing research focuses on refining these methods, particularly in the context of PDAC, to develop robust classification models (such as the TLB score) that can distinguish between healthy individuals and patients with PDAC. The ultimate goal is to improve early diagnosis and assess treatment efficacy using a minimally invasive technique, making TLB a promising tool in the realm of precision medicine.

In advanced diagnostic research, machine-learning (ML) algorithms are employed to analyze complex biomedical data for distinguishing between healthy and diseased states.[39,40] These algorithms can effectively process vast datasets, such as genomics, proteomics, and complex imaging results, enabling the identification of subtle biomarkers and patterns indicative of specific diseases. In the context of conditions where early detection is paramount, ML algorithms can significantly improve diagnostic accuracy, facilitating early intervention and personalized treatment plans. This approach represents a transformative step in precision medicine, leveraging computational power to interpret intricate biological data.

Therefore, the objective of this study was to propose a new methodology for analyzing thermogram curves of biological samples, in this case peripheral blood serum samples. This methodology uses ML tools to get direct information from the DSC curves and to obtain a robust classification model that differentiates between healthy control patients and patients with PDAC. We named it the intelligent TLB [iTLB] model.

2. Experimental Section

2.1. Subjects

For developing a new methodology for thermogram analysis and creating a patient classification model (iTLB model), retrospective serum samples from a discovery cohort, separated into two collections at Blood and Tissue Bank of Aragon and provided by the Biobank of the Aragon Health System, integrated in the Spanish National Biobanks Network, were used: a group of healthy control subjects (blood donors from Zaragoza, $n = 184$) and a group of patients diagnosed with PDAC in the University Miguel Servet Hospital of Zaragoza ($n = 212$). The PDAC group provided clinical information including AJCC eighth edition staging classification indexes (stages 0–IV) and

resectability status (resectable, non-resectable, locally advanced, or metastatic).

Then, the iTLB model was applied to an external and independent cohort of 113 PDAC patients from the University Clinical Hospital of Zaragoza (validation cohort).

To provide a comprehensive comparison of the populations studied, we present the demographic and clinical characteristics of both the training and validation cohorts in **Table 1**.

Serum samples from PDAC group were obtained after the diagnosis and before starting the treatment.

2.2. Thermograms Acquisition

An amount of 5 mL of peripheral blood were collected from each patient into separator gel tubes (BD Vacutainer). The samples were allowed to coagulate before centrifugation at 3000 rpm for 10 min. The serum was aliquoted and stored at −80 °C for later analysis. To ensure patient anonymity, each sample was labeled with an internal code, adhering to the protocol approved by the Ethics Committee.

Thermograms from serum samples were obtained using a high-sensitivity automated capillary VP-DSC differential scanning calorimeter (MicroCal, Malvern-Panalytical). A thermogram consisted of the sample excess heat capacity, C_P, compared to a reference solution as a function of temperature, $C_P(T)$, during thermal denaturation. Serum samples were diluted (1:25) in filtered phosphate-buffered saline and 400 μL was used for testing. The experiments were conducted at a scan rate of $1\,°\text{C}\,\text{min}^{-1}$ from 10 to 95 °C. In case of not being possible to obtain a thermogram from a serum sample, this patient was excluded from the study. In other words, a complete-case analysis was applied without imputation values.

Table 1. Demographic and clinical characteristics of the training and validation cohorts. The table compares key characteristics between the two cohorts, including age distribution, gender, tumor stage, and resectability status. This comparison underscores the similarity between the cohorts, which is crucial for the validation of the iTLB model.

Characteristic	Training cohort ($n = 396$)	Validation cohort ($n = 113$)
Age, years (median, interquartile range)	58 [48; 68] (n.a. = 4)	68 [60; 73] (n.a. = 29)
Sex, n [%]	(n.a. = 1)	(n.a. = 29)
Male	257 (65.06%)	45 (53.57%)
Female	138 (34.94%)	39 (46.43%)
Tumor Stage, n [%]	(n.a. = 46)	(n.a. = 29)
Stage I	0 (0.00%)	15 (17.86%)
Stage II	14 (8.43%)	19 (22.62%)
Stage III	13 (7.83%)	10 (11.90%)
Stage IV	139 (83.74%)	40 (47.62%)
Resectability status, n [%]:	(n.a. = 6)	(n.a. = 29)
Resectable	21 (10.19%)	21 (25.00%)
Locally advanced	46 (22.33%)	23 (27.38%)
Metastatic	139 (67.48%)	40 (47.62%)

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

Thermograms were processed using software developed by the group and implemented in Origin 7 (OriginLab), which involved baseline subtraction and correction, interpolation to get evenly distributed data points in the temperature range ($\Delta T = 0.25\,°C$) and restricting the analysis to the 40–95 °C interval. Each curve was normalized by area before the detailed analysis of changes in the thermograms and their correlation with clinical data.

2.3. Methodology for Developing a Classification Model (TLB Model) from Thermograms

To reduce the number of predictors and avoid overfitting, the information for each degree of temperature was selected for temperatures ranging from 60 to 80 °C. This range was chosen as a target for the ML algorithm, chosen based on prior studies indicating major thermogram changes within this range. This resulted in 21 temperature/predictor variables.

The analysis methodology relied on creating a classification model based on thermogram shapes. The K-Top Scoring Pair (KTSP) function (from R *switchBox* library)[41] used temperature pairs as predictors, an ideal tool for high-dimensional data, reducing overfitting risks and facilitating biological interpretation. The *switchBox* package was applied to convert C_P values into categorical data. This was possible because the algorithm selected a pair of temperatures (T_2, T_1) when there was a strong correlation between the sign of the C_P difference between that pair of temperatures in a group, $C_P(T_2) > C_P(T_1)$ or $C_P(T_2) < C_P(T_1)$, and a correlation of the opposite sign in the other group. For this reason, every thermogram from each patient could achieve the condition of the temperature pairs or not (1/0). Then, a logistic regularized regression applying *ncvreg* package in R, employing Lasso or non-convex penalties and cross validation, was used for model adjustment, trained on 70% of the data, and tested on the remaining 30%, with 100-fold randomization.

Briefly, the methodological procedure consisted of the following steps: 1) creating a resampling matrix to split data 100 times into 70% training data and 30% test data; because the response variable is PDAC or non-PDAC (healthy control group), it is important that the sampling was stratified, that is, maintaining the percentage observed in the initial cohort; 2) determining the optimal range for the number of temperature pairs needed to be compared in the model; 3) selecting the most frequently picked temperature pairs from the 100 fittings; 4) training the model using the most frequently selected temperature pairs; and 5) choosing the best-performing model from the 100 trained models, based on the largest area under the curve (AUC) in the test group.

The outcome of the classification model was a numerical real value per patient within ($-\infty$, $+\infty$), with zero as the standard cutoff. Values below zero classified a patient as a healthy control, and values above zero indicated a patient with pathology (PDAC). The model's performance was evaluated using common indexes from a confusion matrix (sensitivity, specificity, positive predictive value, and negative predictive value), receiver operating characteristic (ROC) curve, and ROC AUC with the 95% confidence interval (CI).

2.4. Statistical Analysis

The study used Kolmogorov–Smirnov–Lilliefors or Shapiro–Wilk tests (based on sample size) to assess variable normality. When appropriate, *t*-tests (unpaired) compared means between two groups, with Bartlett's test used to verify homoscedasticity for normally distributed variables. Wilcoxon's test was employed to compare medians for non-normally distributed variables. A *p*-value <0.05 was significant. Statistical analyses were performed with R version 4.3.2 (October 31, 2023).

3. Results

3.1. Training and Test of iTLB Classification Model: Discriminating between Healthy Subjects and PDAC Patients

To develop the iTLB model, serum samples from 396 individuals were analyzed, comprising two independent groups: a control group of 184 (46.46%) healthy blood donors and a second group consisting of 212 (53.54%) patients diagnosed with PDAC. The majority of the PDAC patients were at stage IV (83.73%, 139/166), while 8.43% (14/166) and 7.83% (13/166) were at stages = II and III, respectively (n.a. = 46) (**Figure** 1A). According to tumor resectability (n.a. = 6), 10.19% (21/206) were resectable, 22.33% (46/206) were locally advanced, and more than half, 67.48% (139/206), were metastatic (Figure 1B).

3.2. iTLB Model Step by Step

For the iTLB classification model based on normalized thermograms of serum samples, 396 serum samples were analyzed: 184 from the healthy control group and 212 from PDAC group. **Figure 2**A,B shows the mean and standard deviation of the thermogram curves (normalized by area) for each group. Figure 2C overlays these means, highlighting the differences between groups. The temperature range 60–80 °C, which contained major thermogram differences, was used for training the classification model.

The automated ML methodology proposed for the analysis of thermograms was applied to obtain a classification model using the thermograms obtained from the 184 healthy subjects and the 212 PDAC patients. First, a resampling matrix was generated to randomize the assign of samples to the training and test groups and repeat this process 100 times. The proportion of patients in each group can be seen in **Figure 3**(1). Second, pairs of temperatures were selected from the 21 predictor temperatures using the KTSP algorithm. The optimal number of pairs was determined by observing the median AUC in the test group (Figure 3(2)). Third, in each training group, temperature pairs were picked, and a classification model was fitted. The pairs selected at least 50 times are shown in Figure 3(3). Fourth, the classification model was trained using only the three most frequent temperature pairs from the last step. These three pairs were picked always (100 times) and are displayed in Figure 3(4). And fifth, at this point, there were 100 trained models, assigning each of them a coefficient to each pair of temperatures. The variability of the coefficients can be visualized by representing the mean plus or minus two times its standard deviation for each of
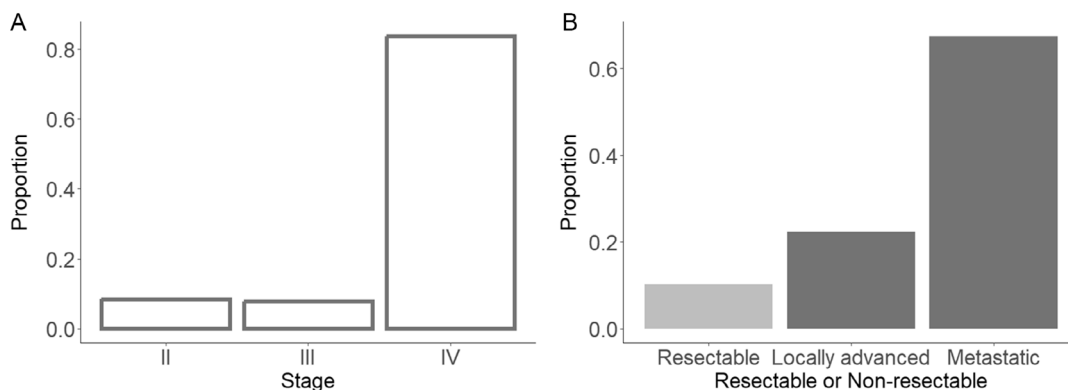
Figure 1. Frequency diagrams of clinical variables in the PDAC patient group. A) Proportion of patients in each stage according to the AJCC eighth edition; B) proportion of patients according to whether the tumor is resectable or not. Stage of the eighth edition of the AJCC (American Joint Committee on Cancer).
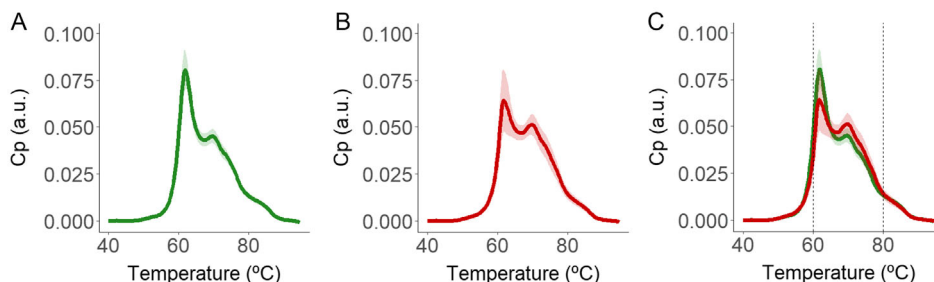

Figure 2. Mean normalized thermogram for each group: A) healthy control subjects in green, and B) PDAC subjects in red, with green/red shading indicating the standard deviation. C) The 60–80 °C temperature range, used for model training, is highlighted with dashed vertical lines. C_P: heat capacity; a.u.: arbitrary units.
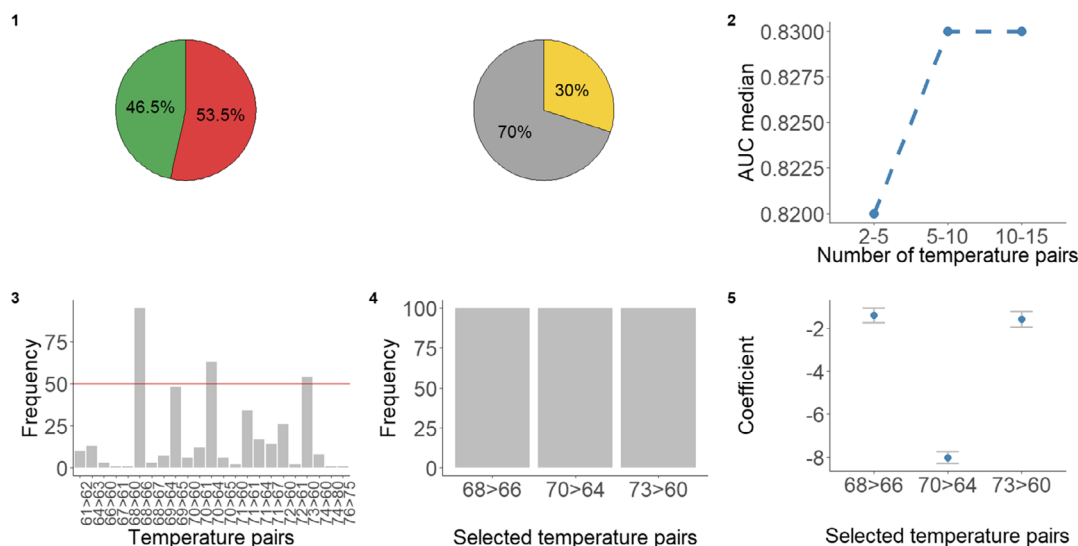

Figure 3. Step-by-step representation of the proposed machine-learning methodology for analyzing thermograms and obtaining a classification model (iTLB model). 1) Left panel: proportion of patients in each group (control subjects in green and PDAC patients in red); right panel: proportion from discovery cohort used for model training and test; 2) median AUC of the 100 test replicates in each temperature pair range. The objective is to determine the optimal number of temperature pairs (elbow method); 3) frequency of each temperature pair's selection in the 100 training replicates when applying the *ncvreg* function. Temperature pairs with a selection frequency higher than 50% over the 100 replicates, which are optimal for the next step; 4) frequency of the selected temperature pairs that passed the 50% cutoff criterion in the previous step in the 100 training replicates; and 5) mean, plus or minus two times the standard deviation, of the coefficients for each temperature pair that makes up the iTLB model.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**

www.advintellsyst.com

the temperature pairs of the iTLB model (Figure 3(5)). Of the 100 trained and validated models, a mean AUC value of 0.83 ($\pm$0.01) was obtained in the training groups and 0.83 ($\pm$0.03) in the test groups.

3.3. iTLB Model

An iTLB model was obtained consisting of only three predictive variables (three pairs of temperatures for which there is a systematic C_P difference in the thermograms from the two groups), where the pair formed by temperatures 64 and 70 °C was clearly the one with the highest statistical weight for differentiating healthy subjects from PDAC patients (**Figure 4**A). Figure 4B shows the location of the three temperature pairs making the iTLB model along the mean thermogram of each group.

The iTLB model provided a numerical value or TLB index for each patient. In the test group, statistically significant differences were observed in the median TLB index between the healthy subject group (-1.03 [$-1.03;0.17$]) and the PDAC patient group (1.46 [0.26;9.66]) (Wilcoxon test: p-value < 0.001) (Figure 4C), with an AUC of 0.90 (95% CI $= 0.85$–0.96) (Figure 4D) in the test group.

To simplify the interpretation of the iTLB model and the TLB index, a cutoff point can be used, which has a standard value of zero. Individuals with TLB index < 0 are classified as healthy subjects, whereas individuals with TLB index > 0 are classified as disease subjects. According to the confusion matrix (Figure 4E), iTLB model had a better sensitivity and

negative predictive value than specificity and positive predictive value.

3.4. Patterns Obtained from the iTLB Model

Since dichotomous predictive variables were used, the TLB index (numerical result from the prediction model) could be considered as a discrete quantitative variable with 2^n possible outcome options (where "n" is the number of predictor variables). Therefore, a maximum of eight results were obtained. By grouping the thermograms of patients with the same TLB index and representing the mean thermograms in each group (control subjects and PDAC patients), several patterns and differences in them could be appreciated (**Figure 5**).

3.5. iTLB Model Applied to an External Cohort of PDAC Patients: Validation of the iTLB Model

An independent, additional cohort of 113 patients diagnosed with PDAC was used to assess the developed iTLB on an external validation group. Thermograms from these 113 PDAC patients were obtained, and the iTLB model was applied to obtain a TLB index for each patient. According to the standard cutoff of zero, if TLB index was greater than zero, the result was considered true positive, but if the TLB index was less than zero, the result was considered false negative. From 113 PDAC patients, 110 were true positive (TLB > 0) and 3 were false negative (TLB < 0). Or in other words, a sensitivity of 97.35% and a false negative rate of 2.65% were obtained.
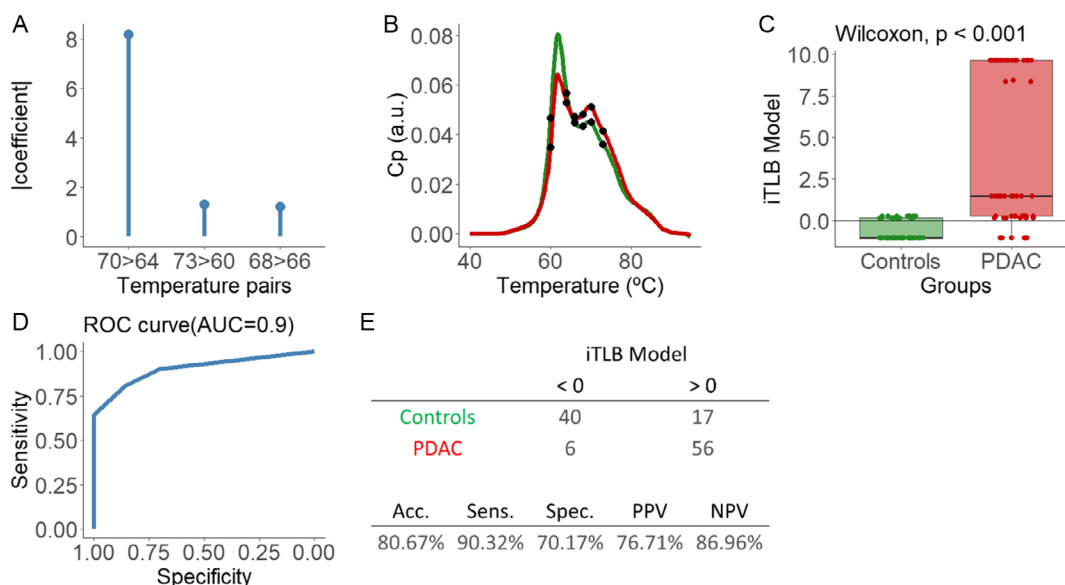

Figure 4. Results of the optimal iTLB model to differentiate the thermograms from healthy control subjects and from PDAC patients. A) Absolute value of the coefficients of the iTLB model in each of the predictor variables (three pairs of temperatures), with $X > Y$ representing each selected pair of temperatures, meaning that the difference $C_P(X) - C_P(Y)$ has systematically a different sign in the thermograms in the PDAC patient group compared to those in the control group; B) means of normalized thermograms in each group (healthy controls and PDAC patients) and the three relevant pairs of temperatures (shown in the x-axis in panel A) in the model are marked with black dots; C) TLB index for the individuals in the test group; D) area under the ROC curve of the iTLB model in the test group; and E) in the top part, contingency table for the test group using the standard cutoff of zero. The bottom part shows the performance indexes of the iTLB model. C_P: heat capacity; a.u.: arbitrary units; iTLB: intelligent thermal liquid biopsy; AUC: area under the ROC curve; Acc: accuracy; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; and NPV: negative predictive value.
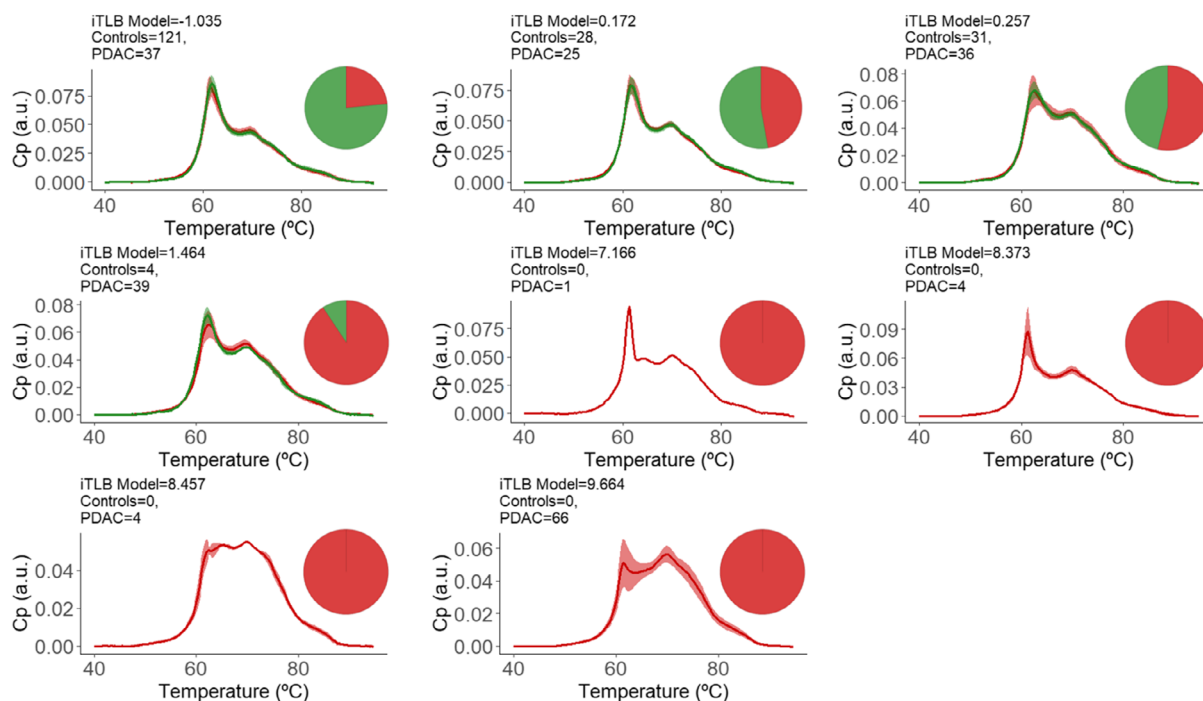
Figure 5. Thermogram patterns obtained after clustering the individuals according to their TLB index. The mean thermograms in each group (control subjects, in green, and PDAC patients, in red) are represented, with shaded area indicating the standard deviation. Each pattern includes the TLB index, the number of control subjects and PDAC patients corresponding to that pattern, and the proportion of control subjects and PDAC patients belonging to each pattern (pie chart). C_P: heat capacity; a.u.: arbitrary units; PDAC: pancreatic ductal adenocarcinoma; and iTLB: intelligent thermal liquid biopsy.

4. Discussion

Pancreatic cancer is one of the deadliest and most aggressive cancers. Patients are often diagnosed in advanced stages due primarily to the lack of early symptoms, the absence of diagnostic tools with sufficient sensitivity and specificity for early-stage detection, and poor response to chemotherapeutic treatments. CA19.9 is the most commonly used biomarker in patients with PDAC, which is more notable for monitoring disease progression as an indicator of tumor burden rather than as a diagnostic tool, due to its low negative predictive value.

Given these challenges, pancreatic cancer is an area where the search for new complementary tools for patient management is particularly relevant. The pursuit of biomarkers in peripheral blood samples is a primary goal, given the ease, speed, and minimal invasiveness of such samples. However, the discovery of new biomarkers has become increasingly difficult. Advances in proteomics and interactome theories suggest that a set of proteins, rather than a single biomarker, may better explain the body's response to a disease. Therefore, new experimental integrative approaches with different biophysical bases should be considered. In this scenario, DSC emerges as a sound option for observing the proteome and peptidome and their interactions.

In our study, the differences observed in the DSC thermograms between PDAC patients and healthy controls may be attributable to several underlying biological mechanisms. The tumor microenvironment in PDAC is known to induce considerable alterations in the serum proteome, reflecting both the systemic inflammatory response and the specific secretion of tumor-related proteins and metabolites. For instance, the presence of elevated levels of acute-phase proteins, such as C-reactive protein and serum amyloid A, as well as changes in the abundance and structure of carrier proteins like albumin, could result in the distinct thermal profiles observed.[42] These proteins, due to their role in the response to malignancy, may form protein complexes or undergo posttranslational modifications,[43] leading to altered thermal stability and denaturation profiles detectable by DSC. Additionally, the release of exosomes and other extracellular vesicles by PDAC cells, which contain a variety of proteins, nucleic acids, and metabolites, could further contribute to the unique thermal signatures seen in the serum of PDAC patients.[44] These mechanisms underscore the potential of DSC as a tool for capturing complex and diverse proteomic alterations associated with PDAC.

The use of DSC with biological samples, i.e., TLB, has shown promise as a potential new complementary tool for diagnosis, monitoring, or prognosis of patients. However, the development of a standardized methodology for analyzing thermograms to differentiate between health/disease states represents a major challenge with TLB. Various groups have contributed with proposals reporting increasingly sophisticated methodologies thanks to bioinformatics advances, facilitating their application and interpretation.

This work reports a new methodology for TLB analysis that is easy to implement, interpret, and reproduce. This methodology

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

employs direct information from thermograms and ML tools, taking advantage of validated algorithms for other types of data, such as omics. The aim is to create a classification model that distinguishes, in this case, between thermograms of healthy control subjects and PDAC patients. The temperature range of 60–80 °C in the thermograms was considered as the main location for relevant changes between both groups, and the C_P value at each temperature (within the 60–80 °C interval) was used as a predictive variable, similar to other groups' approaches.[34] The novel aspect of this methodology for thermogram data analysis is the search for a reduced set of specific temperature pairs containing the relevant information for discriminating between thermograms using the KTSP function. The model was adjusted using the *ncvreg* function, which selects relevant temperature pairs applying cross validation and an embedded classification algorithm with penalization. The use of penalized classification tools for thermogram data has been advocated to avoid overfitting.[35]

The methodology comprises five simple steps, with some subjectivity degree when selecting the most frequent pairs standing out as predictive elements after many replicates within a randomized scheme (step 3). It is proposed to test several model options to achieve an optimal model without overfitting. The methodology was applied to a balanced population of 396 patients, divided into 184 healthy individuals and 212 PDAC patients, to obtain an iTLB classification model. Most PDAC patients were in advanced disease stages. Although this reflects the reality in pancreatic cancer diagnosis, it is advisable to expand the sample size and include more early-stage patients in a future work.

The TLB model developed consists of three pairs of temperatures that provide the relevant information for subject classification. Thus, the data analysis of the complete thermogram is reduced to compare the value of C_P at these three temperature pairs and it provides a unique real value, the TLB index. The model achieved a median AUC of 0.83 ($\pm$0.01) in the training groups, and 0.83 ($\pm$0.03) in the test groups, with significant differences between the TLB index in the healthy subjects group and the PDAC patients group. When training a model, balanced groups are recommended, but if not feasible, balanced training is essential. An unbalanced trained classification model is when one of the groups is more represented than the other category. This leads to biased model training to the detriment of the minority group, which usually contains the cases of greatest interest. For this reason, a balanced training is important to avoid these bias problems.

The interpretation of the TLB index is straightforward, consisting of a single value, for which a cutoff point can be applied to classify subjects. The standard cutoff is zero, so that negative TLB indexes correspond to individuals classified as healthy, and positive TLB indexes correspond to individuals classified as diseased (in this case, with PDAC). The sensitivity and specificity of the iTLB model are comparable to those for CA19.9, but with higher negative predictive value, around 80% for the iTLB model compared to 58% for CA19.9.

Since the iTLB classification model is based on dichotomous predictive variables, responses can be considered discrete quantitative. Under that perspective, different patterns were observed when comparing and grouping thermograms of individuals with the same TLB index. Specific profiles correspond to TLB indexes

exclusive to PDAC patients and are not found in healthy individuals. This "supervised clustering" is a novel approach in relation to thermogram results using biological samples.

In recent years, there has been growing interest in the use of nanoparticle–protein corona characterization as a novel approach for early PDAC detection. These methods involve the analysis of the protein corona that forms, through specific or unspecific interactions, when nanoparticles are incubated with biological fluids, such as serum. The composition of this corona may reflect the pathological state of an individual and serve as a highly sensitive diagnostic marker. For instance, recent studies have developed the nanoparticle-enabled blood test, providing a risk score based on the characterization of the protein corona, which shows potential as a first-level screening tool for PDAC.[45–48] Along our own research, we have explored the use of nanoparticles incubated with serum to detect different signals indicative of disease presence, as demonstrated in our recent publication where we combined TLB and fluorescence spectroscopy to detect serum composition changes following bowel preparation for colonoscopy.[49] Building on this, our ongoing work, which we plan to submit for publication shortly, introduces artificial intelligence (AI) tools for analyzing these complex datasets, further enhancing the diagnostic potential of these methods. Given these advancements, it would be intriguing to explore whether the DSC-based thermogram analysis could be integrated with nanoparticle–protein corona characterization and AI-driven data analysis. Such an integrated approach could enhance the sensitivity and specificity of PDAC detection by capturing multiple facets of the proteomic landscape associated with the disease.

The iTLB model was applied to an external cohort of 113 PDAC patients like validation cohort, where it was obtained a sensitivity similar to that of the discovery cohort. This result confirms the robustness of the iTLB model in PDAC patients. It is necessary to have an external healthy group to validate the iTLB model. The next step is to further apply this methodology to a larger sample to confirm the applicability of TLB in the diagnosis of PDAC using a group symptomatic patients without cancer or with pancreatic pathologies nonmalignant as a control group.

As a conclusion, a new methodology for analyzing thermograms of biological samples is proposed, using serum samples from healthy controls and PDAC patients as an example, to obtain a robust iTLB model based on ML tools that differentiates between these two groups. The results of the iTLB model demonstrate diagnostic capability, in terms of sensitivity and specificity, comparable to the tumor marker CA19.9, which is currently used in clinical settings. However, the iTLB model shows a higher negative predictive value. This finding opens the possibility of using the iTLB model as a complementary tool to improve the diagnosis of pancreatic cancer patients.

While the DSC-based methodology for early PDAC diagnosis shows considerable promise for diagnostic and patient monitoring, there are several limitations that should be acknowledged. First, the requirement for specialized equipment, such as a high-sensitivity DSC instrument, may limit the widespread adoption of this technique in routine clinical settings, particularly in resource-limited environments. Additionally, the interpretation of DSC thermograms necessitates expertise in both biophysical methods data analysis and ML tools, which may not be readily available in all clinical laboratories. Moreover, the current study

focused on the application of this methodology to PDAC; however, further research is needed to evaluate its applicability across other types of cancers and diseases, as well as in different scenarios and contexts. Another limitation is the retrospective nature of the study, which, although mitigated by the use of an independent validation cohort, still necessitates prospective studies to confirm the utility of this approach in a real-world clinical setting. Finally, while our method shows considerably high sensitivity, the potential for false positives, particularly in patients with inflammatory or benign pancreatic conditions, needs to be carefully evaluated in future studies to ensure the specificity of the approach.

Further studies are necessary to confirm these findings. The methodology should also be applied to other diseases and types of biological samples. It is also important to demonstrate that the incorporation of clinically used biomarkers into the iTLB model might improve the diagnosis, prognosis, or monitoring of some diseases. This approach may represent a significant advancement in personalized medicine, offering more accurate and tailored treatment strategies for patients.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Sonia Hermoso-Duran: Conceptualization (equal); Data Curation (Equal); Formal Analysis (Equal); Investigation (Equal); Visualization (Equal); Writing—Original Draft (Equal); and Writing—Review and Editing (Equal). **Nicolas Fraunhoffer**: Data Curation (Equal); Formal Analysis (Equal); Methodology (Equal); Writing—original draft (equal); and Writing—Review and Editing (equal). **Judith Millastre-Bocos**: Investigation (Equal); Resources (Equal); Writing—Original Draft (Supporting); and Writing—Review and Editing (Supporting). **Oscar Sanchez-Gracia**: Formal Analysis (Equal); Methodology (Supporting); Software (Equal); and Validation (Supporting). **Pablo F. Garrido**: Data Curation (Supporting); Formal Analysis (Supporting); Methodology (Supporting); and Software (Supporting); Validation (Supporting). **Sonia Vega**: Data Curation (Equal) and Investigation (Equal). **Ángel Lanas**: Project Administration (Supporting); Supervision (Supporting); and Writing—Review and Editing (Supporting). **Juan Iovanna**: Project Administration (Supporting); Supervision (Supporting); and Writing—Review and Editing (Supporting). **Adrian Velazquez-Campoy**: Conceptualization (Equal); Funding Acquisition (Equal); Investigation (Equal); Supervision (Equal); Writing—Original Draft (Equal); and Writing—Review and Editing (Equal). **Olga Abian**: Conceptualization (Equal); Funding Acquisition (Equal); Project Administration (Equal); Resources (Equal); Supervision (Equal); Validation (Equal); Writing—Original Draft (Equal); and Writing—Review and Editing (Equal).

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

[1] O. Partyka, M. Pajewska, D. Kwaśniewska, A. Czerw, A. Deptała, M. Budzik, E.ż. Cipora, I. Gąska, L. Gazdowicz, A. Mielnik, K. Sygit, M. Sygit, E. Krzych-Fałta, D. Schneider-Matyka, S. Grochans, A. M. Cybulska, J. Drobnik, E. Bandurska, W. Ciećko, P. Ratajczak, K. Kamecka, M. Marczak, R. Kozłowski, *Cancers* **2023**, *15*, 3634.

[2] M. Korc, C. Y. Jeon, M. Edderkaoui, S. J. Pandol, M. S. Petrov, *Best Pract. Res. Clin. Gastroenterol.* **2017**, *31*, 529.

[3] G. Lippi, C. Mattiuzzi, *Arch. Med. Sci.* **2020**, *16*, 820.

[4] G. Gheorghe, S. Bungau, M. Ilie, T. Behl, C. M. Vesa, C. Brisc, N. Bacalbasa, V. Turi, R. S. Costache, C. C. Diaconu, *Diagnostics* **2020**, *10*, 869.

[5] A. Di Federico, V. Tateo, C. Parisi, F. Formica, R. Carloni, G. Frega, A. Rizzo, D. Ricci, M. Di Marco, A. Palloni, G. Brandi, *Pharmaceuticals* **2021**, *14*, 677.

[6] D. Casado, R. V. Tocino, E. Fonseca, B. Cigarral, B. Barrios, E. Escalera, J. Claros, L. Figuero, A. Olivares, Á. López, E. Terán, J. J. Cruz, *Med. - Programa de Formación Médica Continuada Acreditado* **2021**, *13*, 1345.

[7] S. C. Lindgaard, Z. Sztupinszki, E. Maag, I. M. Chen, A. Z. Johansen, B. V. Jensen, S. E. Bojesen, D. L. Nielsen, C. P. Hansen, J. P. Hasselby, K. R. Nielsen, Z. Szallasi, J. S. Johansen, *Clin. Cancer Res.* **2021**, *27*, 2592.

[8] L. D. Mellby, A. P. Nyberg, J. S. Johansen, C. Wingren, B. G. Nordestgaard, S. E. Bojesen, B. L. Mitchell, B. C. Sheppard, R. C. Sears, C. A. K. Borrebaeck, *J. Clin. Oncol.* **2018**, *36*, 2887.

[9] M. B. Amin, S. B. Edge, F. L. Greene, R. L. Schilsky, L. E. Gaspar, M. K. Washington, et al., *AJCC Cancer Staging Manual*, 8th ed., Springer International Publishing, New York **2017**, https://doi.org/10.1007/978-3-319-40618-3.

[10] N. C. Garbett, M. L. Merchant, C. W. Helm, A. B. Jenson, J. B. Klein, J. B. Chaires, J. M. Sanchez-Ruiz, *PLoS One* **2014**, *9*, e84710.

[11] N. C. Garbett, C. S. Mekmaysy, C. W. Helm, A. B. Jenson, J. B. Chaires, *Exp. Mol. Pathol.* **2009**, *86*, 186.

[12] N. C. Garbett, J. J. Miller, A. B. Jenson, J. B. Chaires, *Semin. Nephrol.* **2007**, *27*, 621.

[13] M. Zhou, D. A. Lucas, K. C. Chan, H. J. Issaq, E. F. Petricoin, L. A. Liotta, T. D. Veenstra, T. P. Conrads, *Electrophoresis* **2004**, *25*, 1289.

[14] M. Vidal, M. E. Cusick, A.-L. Barabási, *Cell* **2011**, *144*, 986.

[15] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási, *Science* **2015**, *347*, 1257601.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

[16] N. C. Garbett, J. J. Miller, A. B. Jenson, J. B. Chaires, *Biophys. J.* **2008**, *94*, 1377.

[17] S. Krumova, S. Todinova, S. G. Taneva, *Cancers* **2022**, *14*, 3884.

[18] S. Todinova, S. Krumova, L. Gartcheva, C. Robeerst, S. G. Taneva, *Anal. Chem.* **2011**, *83*, 7992.

[19] I. Zapf, T. Fekecs, A. Ferencz, G. Tizedes, G. Pavlovics, E. Kálmán, D. Lőrinczy, *Thermochim. Acta* **2011**, *524*, 88.

[20] Z. Szalai, T. F. Molnár, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2013**, *113*, 259.

[21] M. Mehdi, T. Fekecs, I. Zapf, A. Ferencz, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2013**, *111*, 1801.

[22] N. C. Garbett, M. L. Merchant, J. B. Chaires, J. B. Klein, *Biochim. Biophys. Acta, Gen. Subj.* **2013**, *1830*, 4675.

[23] M. Moezzi, A. Ferencz, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2014**, *116*, 557.

[24] N. A. Kim, J. H. Jin, K.-H. Kim, D. G. Lim, H. Cheong, Y. H. Kim, W. Ju, S. C. Kim, S. H. Jeong, *Arch. Pharmacal Res.* **2016**, *39*, 668.

[25] I. Zapf, M. Moezzi, T. Fekecs, K. Nedvig, D. Lőrinczy, A. Ferencz, *J. Therm. Anal. Calorim.* **2016**, *123*, 2029.

[26] A. Ferencz, I. Zapf, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2016**, *126*, 55.

[27] Z. Szalai, T. F. Molnár, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2017**, *127*, 1231.

[28] A. Ferencz, D. Lőrinczy, *J. Therm. Anal. Calorim.* **2017**, *127*, 1187.

[29] A. Michnik, E. Sadowska-Krępa, J. Cholewa, I. Schisler, A. Kiełboń, Z. Drzazga, *Thermochim. Acta* **2018**, *662*, 64.

[30] D. Lőrinczy, M. Moezzi, A. Ferencz, *J. Therm. Anal. Calorim.* **2020**, *142*, 789.

[31] D. Lőrinczy, A. Ferencz, *J. Therm. Anal. Calorim.* **2020**, *142*, 1243.

[32] S. Brudar, U. Černigoj, H. Podgornik, M. Kržan, I. Prislan, *Acta Chim. Slov.* **2017**, *64*, 564.

[33] D. J. Fish, G. P. Brewood, J. S. Kim, N. C. Garbett, J. B. Chaires, A. S. Benight, *Biophys. Chem.* **2010**, *152*, 184.

[34] S. Rai, S. Pan, S. Cambon, S. Chaires, S. Garbett, *Open Access Med. Stat.* **2013**, *3*, 1.

[35] N. C. Garbett, G. N. Brock, *Biochim. Biophys. Acta, Gen. Subj.* **2016**, *1860*, 981.

[36] S. K. Kendrick, Q. Zheng, N. C. Garbett, G. N. Brock, Y. Barenholz, *PLoS One* **2017**, *12*, 0186232.

[37] N. C. Garbett, G. N. Brock, J. B. Chaires, C. S. Mekmaysy, L. Deleeuw, K. L. Sivils, J. B. Harley, B. H. Rovin, K. B. Kulasekera, W. N. Jarjour, X. Zhou, *PLoS One* **2017**, *12*, 0186398.

[38] S. N. Rai, S. Srivastava, J. Pan, X. Wu, S. P. Rai, C. S. Mekmaysy, L. Deleeuw, J. B. Chaires, N. C. Garbett, J. M. Sanchez-Ruiz, *PLoS One* **2019**, *14*, 0220765.

[39] N. Caballé-Cervigón, J. L. Castillo-Sequera, J. A. Gómez-Pulido, J. M. Gómez-Pulido, M. L. Polo-Luque, *Appl. Sci.* **2020**, *10*, 5135.

[40] P. Singh, N. Singh, K. K. Singh, A. Singh, *Machine Learning and the Internet of Medical Things in Healthcare*, Elsevier, London **2021**, pp. 89–111, https://doi.org/10.1016/B978-0-12-821229-5.00003-3.

[41] B. Afsari, E. J. Fertig, D. Geman, L. Marchionni, *Bioinformatics* **2015**, *31*, 273.

[42] D. Caputo, A. Coppola, E. Quagliarini, R. Di Santo, A. L. Capriotti, R. Cammarata, A. Laganà, M. Papi, L. Digiacomo, R. Coppola, D. Pozzi, G. Caracciolo, *Cancers* **2022**, *14*, 4658.

[43] S. Pan, R. Chen, *Mol. Aspects Med.* **2022**, *86*, 101097.

[44] A. Nicoletti, M. Negri, M. Paratore, F. Vitale, M. E. Ainora, E. C. Nista, A. Gasbarrini, M. A. Zocco, L. Zileri Dal Verme, *Int. J. Mol. Sci.* **2023**, *24*, 885.

[45] R. Di Santo, L. Digiacomo, E. Quagliarini, A. L. Capriotti, A. Laganà, R. Zenezini Chiozzi, D. Caputo, C. Cascone, R. Coppola, D. Pozzi, G. Caracciolo, *Front. Bioeng. Biotechnol.* **2020**, *8*, 491.

[46] M. Papi, G. Caracciolo, *Nano Today* **2018**, *21*, 14.

[47] M. Papi, V. Palmieri, L. Digiacomo, F. Giulimondi, S. Palchetti, G. Ciasca, G. Perini, D. Caputo, M. C. Cartillone, C. Cascone, R. Coppola, A. L. Capriotti, A. Laganà, D. Pozzi, G. Caracciolo, *Nanoscale* **2019**, *11*, 15339.

[48] L. Digiacomo, E. Quagliarini, D. Pozzi, R. Coppola, G. Caracciolo, D. Caputo, *Cancers* **2023**, *15*, 2983.

[49] S. Hermoso-Durán, M. J. Domper-Arnal, P. Roncales, S. Vega, O. Sanchez-Gracia, J. L. Ojeda, Á. Lanas, A. Velazquez-Campoy, O. Abian, *Cancers* **2023**, *15*, 1952.