Particle Swarm Optimized Gaussian Process Classifier for Treatment Discontinuation Prediction in Multicohort Metastatic Castration-Resistant Prostate Cancer Patients

Olutomilayo Olayemi Petinrin, Xiangtao Li, and Ka-Chun Wong

***Abstract*—Prostate cancer is the second leading cancer in men, according to the WHO world cancer report. Its prevention and treatment demand proper attention. Despite numerous attempts for disease prevention, prostate tumours can still become metastatic by blood circulation to other organs. Several treatments have been adopted. However, findings show that the docetaxel treatment induces adverse reactions in patients. Particle Swarm Optimized Gaussian Process Classifier (PSO-GPC) is proposed to determine when to discontinue treatment. Based on three cohorts of prostate cancer patients, we propose and compare several classifiers for the best performance in determining treatment discontinuation. Given the data skewness and class imbalance, the models are evaluated based on both the area under receiver operating characteristics curve (AUC) and area under precision recall curve (AUPRC). With the AUCs ranging between 0.6717-0.8499, and AUPRCs ranging between 0.1392-0.5423, PSO-GPC performs better than the state-of-the-art. We have carried out statistical analysis for ranking methods and analyzed independent cohort data with PSO-GPC, demonstrating its unbiased performance. A proper determination of treatment discontinuation in metastatic castration-resistant prostate cancer patients will reduce the mortality rate in cancer patients.**

***Index Terms*—Gaussian process classifier, machine learning, metastatic castration-resistant prostate cancer, multi-cohort data, particle swarm optimization.**

Manuscript received March 5, 2021; revised July 14, 2021; accepted August 3, 2021. Date of publication August 11, 2021; date of current version March 7, 2022. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region under Grant CityU 11200218, in part by the Health and Medical Research Fund, the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region under Grant 07181426, in part by the Hong Kong Institute for Data Science (HKIDS) at City University of Hong Kong, in part by the City University of Hong Kong under Grants CityU 11202219 and CityU 11203520, in part by the National Natural Science Foundation of China research project under Grant 32000464, and in part by the Shenzhen Research Institute, City University of Hong Kong. *(Corresponding author: Ka-Chun Wong.)*

Olutomilayo Olayemi Petinrin and Xiangtao Li are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong SAR (e-mail: opetinrin2-c@my.cityu.edu.hk; lixt314@jlu.edu.cn).

Ka-Chun Wong is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong SAR, and also with Hong Kong Institute for Data Science, City University of Hong Kong, Kowloon, Tong, Hong Kong SAR (e-mail: kc.w@cityu.edu.hk).

Digital Object Identifier 10.1109/JBHI.2021.3103989

I. Introduction

With the continued surge in the number of people affected by cancer, the earlier its detection, the higher the chances of living [1]. Prostate cancer, a form of cancer peculiar to men grows in the prostate gland. It is the second leading death-causing cancer in men in the USA [2]. Popular radiation therapy and chemotherapy have been challenging and ineffective for treating tumor metastasis, which has made it responsible for more than 90% of deaths related to cancer. It is an open problem in cancer biology [3]. Tumor metastasis aims for highly belligerent tumor cells with the tendency of dynamism and ability to grow in distant tissue microenvironments. It contributes to the high mortality associated with cancer patients. Factors such as oxygen, which aids oxidative metabolism, the survival of cells, and adenosine 5'-triphosphate (ATP) generation, influence the growth of these tumors in the primary location [4].

The consequence of metastasis treatment can sometimes contribute to resistance to drugs and rapid mortality from the disease [5]. In some cases, recognizing the sensitivities of some medications to different forms of existing cancers might be difficult. With metastasis treatment being a challenging area in medicine, extensive medical studies have been conducted to determine the best solution and to ensure quality-life longevity [6], [7].

Deprivation of androgen, also known as Androgen Depletion Therapy (ADT), is expected to decrease testosterone production in the male human body, which would also control the growth and spread of cancer cells. In some cases, although a seemingly initial response is noticed, the cell growth is later uncontrollable, which leads to a state known as Metastatic Castration-Resistant Prostate Cancer (mCRPC) [8], [9]. Considering the risk level, a first-hand chemotherapy treatment approach for metastatic cases is recommended for increased chances of survival [10]. Detection has to be made early to prompt treatment. However, it sometimes contributes to overtreatment, resulting in side effects

that may reduce life quality. With the advent of new drugs [11]–[13] in addition to docetaxel, improved choices can be made for the treatment of mCPRC; nonetheless, with clinical treatment options such as chemotherapy and hormone therapy, the optimal method to prevent treatment resistance remains uncertain [14].

Despite the beneficial results docetaxel has exhibited in treating patients with mCRPC, the management of the risk factors that arise due to toxicity-related adverse effect has been a central challenge [15]. It is necessary that optimal decisions are made for each patient while balancing the treatment to prevent overtreatment and progression of the disease [16]. The study of different drugs and treatments relating to mCPRC is largely influenced by its mortality-causing adverse effects.

In recent times, machine learning techniques have been applied to uncover new truths and make predictions to enhance the detection of prostate cancer, especially in medical images. As a powerful tool, machine learning methods can be applied to several aspects of prostate cancer study, such as detection, localization, and assessment of aggression [17]. It has shown better accuracy than certified radiologists in predicting Gleason prostate cancers [18], [19]. In addition, machine learning can be used to effectively predict docetaxel discontinuation in patient with metastatic castration-resistant prostate cancer [20].

The metastatic state of cancer is a challenging area in cancer research, ranging from the investigation of microenvironment to treatment options. Therefore, the search for prevention and remedies are always desired. The determination of treatment discontinuation in metastatic castration-resistant prostate cancer patients can contribute to the reduction of toxicity-induced mortality and hence improve life quality [21]. Based on particle swarm optimization for the Gaussian process classifier, we present a prediction model and compare its performance with standard classifiers and existing related works. Due to the non-parametric property of the Gaussian process and its successful application in regression, its less-explored prospect for classification is utilized in this study. The Gaussian process classifier is optimized with the particle swarm optimization method for its practical and robust convergence capability [22], [23].

II. Methods

A. Dataset Composition

The comparator arm of a prostate cancer dataset was collected, which is a combination of three cohorts (Venice, Ascent-2, and Celgene) with the information of metastatic-castration resistant prostate cancer patients. These data can be accessed from the Project Data Sphere Cancer Research Platform [24]. Information about Ascent-2, Celgene, Venice cohorts and the entire cohort combination (hereafter referred to as All cohort) is given in Table I. The datasets were split into train and test set in a 70 to 30 ratio, and 5-fold cross validation was implemented with grid search during training for the selection of parameters to prevent overfitting. Subsequently, we later set Celgene cohort apart as an independent test set while Ascent-2 and Venice cohort were combined as training data.

TABLE I
Detailed Information About Datasets

Dataset Reg. No	Data Provider	Number of Sample	Training (70%)	Test (30%)	Discont. Rate
Ascent-2 NCT 00273338	Memorial Sloan Kettering Cancer Center	476	333	143	22.06%
Celgene NCT 00988208	Celgene Corporation	526	368	158	7.79%
Venice NCT 00519285	Sanofi US Services Inc.	598	418	180	5.53%
All		1600	1120	480	12.32%

B. Data Preprocessing

According to the data dictionary associated with the dataset, the data contains 131 features, some of which represent basic information about the data source, metastasis tumor locations and medical history.

1) Data Imputation: Missing values has quite an impact on predictive analysis and the generated result [25] especially in clinical trials [26], [27]. This makes data imputation important. The original dataset had quite some missing values. In some binary features, the columns only contain "Yes". However, according to the data dictionary, the missing areas should be imputed with "No". Other instances of missing values also exist which can have huge impacts on the analysis of the entire dataset. Due to the size of the instances with missing values, those instances cannot simply be discarded. For the attributes with numerical values, the missing values were replaced with the mean $\bar{x} = \frac{\sum x_i}{N}$ of the available values; for the attributes with nominal values, the missing values were replaced with the mode of the available values. Other features(such as cohort name, and patient's IDs,) considered irrelevant to the analysis were removed.

2) Data Encoding and Normalization: We split the columns according to the data types and applied one-hot encoding to the attributes with categorical values to handle the categories' binarization and prevent the assumption of ordinal relationships. Attributes with numerical values were normalized as z-scores.

3) Kernel Principal Component Analysis: The classification of nonlinear data is non-trivial because the hyperplanes are not easily defined as in the case of linear data. The standard Principal Component Analysis (PCA) is best suited for linearly separable data, while the kernel PCA is implemented for dimensionality reduction of linearly non-separable data. This technique handles multicollinearity, which affects variance in the data. Using Radial Basis Function (RBF) kernel and transforming the input data to a high-dimensional space where it is linearly separated, the mapping can be written as $x \rightarrow \phi(x)$ given the nonlinear mapping function ϕ, for each data sample x. The covariance matrix based kernel trick, which offers an efficient and computational cost-effective way of transforming data to a higher dimension, was applied to obtain the eigenvectors α, by which the data is projected on the principal components.

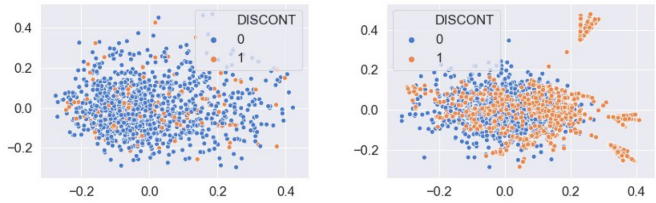

Fig. 1. 2D visualization of class distribution before and after oversampling.

It enables functionality in the original feature space without calculating the data coordinates in a higher dimensional space.

$$\text{Covariance} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T \tag{1}$$

For every pair of data point, the similarity was calculated based on:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|_2^2) \tag{2}$$

This resulted in an $N \times N$ symmetric kernel matrix. To center the symmetric kernel matrix, (3) was applied

$$K' = K - 1_N K - K 1_N + 1_N K 1_N \tag{3}$$

where 1_N denotes the matrix $N \times N$ of 1 s divided by N. At this point, to select the principal components which maximize the variance in the dataset, the eigenvectors of the newly centered kernel matrix corresponding to the largest eigenvalues were obtained based on (1). We used 100 principal components in this study.

To transform the test data based on the train data, we project the new data point x onto the principal component axis g by computing $\phi x^T g$. We calculate the kernel between the new data points and every data point j in the training data:

$$\phi x^T g = \sum_j \alpha_i \phi(x)\phi(x_j)^T = \sum_j \alpha_i k(x, x_j) \tag{4}$$

such that the eigenvalues λ and eigenvectors α of the kernel matrix K satisfy $K\alpha = \lambda\alpha$ [28].

4) Oversampling: A major challenge with imbalanced data lies in the overestimated performance on the majority classes. In this study, we used Borderline-SMOTE SVM [29] as the oversampling technique [30]. This technique generates synthetic minority class samples near the decision boundary as shown in Fig. 1. The newly synthesized data points increase the occurrence of the minority class samples for training purpose.

C. Particle Swarm Optimized Gaussian Process Classifier (PSO-GPC)

The Gaussian process, used for regression or classification task, is a non-parametric method which places a Gaussian distribution over unknown functions. We adopt binary classification as it is the focus of our study. The classification principle is based on the prior smoothness while ensuring a good fit for the observed data. Given a set of N training input points $X = [x_1, \ldots, x_N]^T$, each with the corresponding class labels $y = [y_1, \ldots, y_N]^T$ where $\mathbf{y} \in \{0, 1\}$, we determine the class of a new data point $\mathbf{x}_*$ based on its posterior probability $p(\mathbf{y}|\mathbf{x})$.

In [31], Gaussian process is specified by a positive definite covariance function $k(\mathbf{x}; x') = \mathrm{V}[f(\mathbf{x}), f(x')]$ and a mean function $m(x) = \mathrm{E}[f(\mathbf{x})]$. By first computing the distribution of the latent variable f_* which corresponds to a new test point $\mathbf{x}_*$ for the posterior $p(f|X, \mathbf{y})$, we have:

$$p(f_*|\mathbf{x}_*, X, \mathbf{y}) = \int p(f_*|\mathbf{x}_*, X, \mathbf{f}) p(\mathbf{f}|X, \mathbf{y}) df \tag{5}$$

where $f = [f_1, \ldots, f_N]^T$. After that, we compute the probabilistic prediction for the new test point x_*

$$p(y_* = 1|\mathbf{x}_*, X, \mathbf{y}) = \int \Phi(f_*) p(f_*|\mathbf{x}_*, X, \mathbf{y}) df_* \tag{6}$$

The non-Gaussian likelihood in (5) makes the integral analytically intractable. Given its analytical intractability, the non-Gaussian joint posterior is approximated as a Gaussian posterior using Laplace analytic approximation method [32], [33]. To derive the Gaussian approximate $q(f|X, y)$ to the posterior in (5), we can derive the prior $p(f|X)$ which is Gaussian $f|X \sim N(0, k)$ as:

$$\log p(f|X) = -\frac{1}{2}\mathbf{f}^T K^{-1}\mathbf{f} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi \tag{7}$$

From (7), K is the correlation function, a function of the training input and the hyperparameters. The correlation function's behaviour is affected by the choice of hyperparameter, and it is desirable to select the values such that it maximizes the likelihood functions in (7).

Therefore, we find the set of parameters (hyperparameters) that optimizes (7). In complex cases, the function is usually multimodal; and this complexity necessitates the use of global non-convex optimization algorithms [34]. Hence for the optimization task of the classifier [35], we propose the use of Particle Swarm Optimization (PSO) algorithm due to its fast and robust convergence algorithm.

PSO, a population-based search algorithm, follows the bird flocking social behavior. It is a stochastic optimization technique similar to evolutionary techniques such as genetic algorithm [36]. For each iteration, each particle, i is updated by its own best value, P_{best}, and the best value of any particle in the population, G_{best}, which is the global best solution in dimension j through time t. In addition, for each particle, its velocity vector $\mathbf{V}$ and position vector $\mathbf{X}$ are updated by (8) and (9), respectively. This update makes the PSO algorithm less dependent on the initial points/particles.

$$\mathbf{V}_{ij}^{t+1} = w\mathbf{V}_{ij}^t + c_1 r_1^t (P_{best}^t - \mathbf{X}_{ij}^t) + c_2 r_2^t (G_{best} - \mathbf{X}_{ij}^t) \tag{8}$$

$$\mathbf{X}_{ij}^{t+1} = \mathbf{X}_{ij}^t + \mathbf{V}_{ij}^{t+1} \tag{9}$$

Where w is the positive inertia weight constant by which the global and local search is balanced (we used 0.5), c_1 and c_2 denote the positive acceleration constants used to level the contribution of the cognitive and social components (we set it to 1 and 2, respectively), and r_1, r_2 represents a random number from uniform distribution $U(0, 1)$ which is generated at every

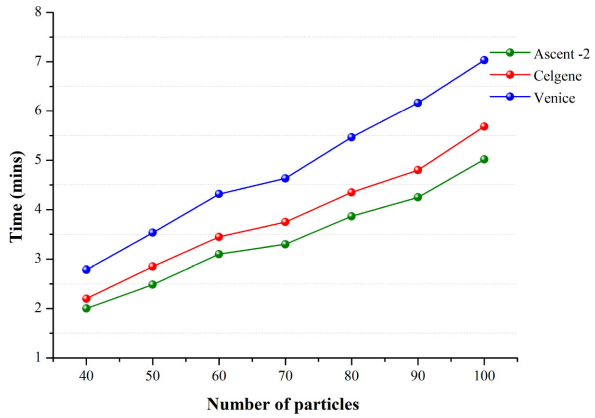

Fig. 2. Runtime of PSO-GPC on Ascent-2, Celgene, and Venice Cohort Datasets.

velocity update. We have used 100 particles and 30 iterations for our study.

With the optimized parameters, we can hence make the prediction using $\int \sigma(f_*)q(f_*|\mathbf{x}_*, X, \mathbf{y})df_*$, where $q(f_*|\mathbf{x}_*, X, \mathbf{y})$ contain the the Gaussian mean and variance derived from the Laplace approximation.

The mechanism of the algorithm is based on the time required to create the initial population and update the solution, making the complexity a function of the problem type and the population size. The PSO algorithm used in the optimization of the classifier in selecting values which maximizes likelihood function is given in Algorithm 1. Its integration into the Gaussian Process Classifier is shown in Algorithm 2. Although the objective function is a strong determinant in the main computational cost, the complexity of the PSO algorithm to check for termination condition is:

$$T(n) = 1 + k + (k.n) + (k.n.logn) + C + 1$$
$$T(n) = k + (k.n) + (k.n.logn) + C$$

In general, the complexity of the algorithm is $O(k.nlogn)$ where k is the number of iterations and n is the number of particles. By maintaining the number of iterations as 30, we successively increased the number of particles from 40 to 100. The resulting runtime for Ascent-2, Celgene, and Venice datasets is as depicted in Fig. 2. We also note that runtime is directly proportional to the number of instances in the datasets. The difference in evaluation score at each particle is minute and in the same range. Since PSO makes updates based on the global best particle, the algorithm halts when a good function is met. This technique makes it easier for convergence to be achieved quickly.

All analyses in this study were carried out on a Windows 10 64-bit Operating System, X64-based processor computer with 16 GB RAM. Processor specification is Intel (R) Core (TM) i7-10510 U CPU @ 1.80 GHz.

D. Evaluation Metrics

Due to the high data imbalance and sparsity, the evaluation of models was based on both the Area Under Receiver Operating Characteristic (ROC) Curve (AUC) and the Area Under

Algorithm 1: PSO Algorithm.

Initialize particles' velocity (V_i), particles' position (X_i), previous best position (P_i), number of particles (N)
while $(k <$ maximum number of iterations $K)$ **do**
  for all particles (i) **do**
    calculate the fitness function for the current position x_i of the ith particle $(F(x_i))$
    if $(F(x_i) < F(P_i))$ **then**
      $P_i = x_i$
    end if
    if $(F(x_i) < F(G))$ **then**
      $G = x_i$
    end if
    Adjust the velocity and positions of all particles according to 8 and 9
  end for
  Stop the algorithm if a sufficiently good function is met.
end while

Algorithm 2: PSO-GPC Algorithm.

Initialize $f = 0$
Compute initial $\nabla\varphi$ and W
where, $\nabla\varphi = K(\nabla logp(y|f)$
and, $W = -\nabla\nabla logp(y|f)$
Using PSO, $min\nabla\varphi$
where, $(\nabla\varphi_{new} = F(f^{new}))$
and, $f^{new} = f - (\nabla\nabla\varphi)^{-1}\nabla\varphi$
$= (K^{-1} + W)^{-1}(Wf + \nabla logP(y|f)$
Compute $logq(y|X, \theta)$
where $logq(y|X, \theta) =$
$-\frac{1}{2}f^T K^{-1} f + \log P(y|f^n) - \frac{1}{2}\log[|K|.|K^{-1} + W|]$
Return $\hat{f} = f^{final}$, $\log q(y|X, \theta)$ (approximated log marginal likelihood)

the Precision-Recall Curve (AUPRC) [37], [38]. The curves determine the model performance without dependence or bias towards the size of the test data used for evaluation. These metrics incorporate both sensitivity and specificity into one metric. For a skewed dataset, accuracy might not be a good evaluation metric as there can be biases towards the majority class. AUC is a measure of overall performance over all possible thresholds. A model with an AUC score close to 1 is considered good. A curve is generated when the True Positive Rate (TPR) is plotted against the False Positive Rate (FPR) based on (10) and (11) and the area under the generated curve is the AUC.

$$\text{TPR} = \frac{(\text{True Positive (TP)})}{(\text{True Positive (TP)+False Negative (FN)})} \quad (10)$$

$$\text{FPR} = \frac{(\text{False Positive (FP)})}{(\text{True Negative (TN)+False Positive (FP)})} \quad (11)$$

AUPRC is an important metric to use when it is essential for the model to correctly predict all positive values while avoiding the prediction of negative values as positive. Especially when

TABLE II
PERFORMANCE OF METHODS ON COHORTS

Methods	Metrics	All	Ascent-2	Celgene	Venice
PSO-GPC	AUC	0.8067	0.8499	0.6717	0.7718
	AUPRC	0.3979	0.5423	0.2420	0.1392
CB	AUC	0.7313	0.7820	0.6066	0.5624
	AUPRC	0.3265	0.4854	0.2151	0.0841
LGB	AUC	0.7895	0.7584	0.6255	0.5947
	AUPRC	0.3898	0.4730	0.2066	0.0802
RF	AUC	0.7877	0.7453	0.5689	0.4974
	AUPRC	0.3830	0.4316	0.1844	0.0743
KNN	AUC	0.6335	0.6718	0.5115	0.5047
	AUPRC	0.2054	0.3242	0.1217	0.0601
XBG	AUC	0.7938	0.7475	0.5000	0.6041
	AUPRC	0.3808	0.4135	0.1203	0.0895

TABLE III
COMPARISON OF CLINICAL IMPORTANCE BETWEEN PSO-GPC AND STATE-OF-ART RESULTS [40]

Cohorts	All	Ascent-2	Celgene	Venice
PSOGPC AUC	**0.8067**	**0.8499**	**0.6717**	**0.7718**
State-of-Art AUC	0.6356	0.5726	0.6420	0.5490
PSOGPC AUPRC	**0.3979**	**0.5423**	**0.2420**	**0.1392**
State-of-Art AUPRC	0.2001	0.3089	0.1598	0.1006
Baseline AUPRC	0.1232	0.2206	0.0779	0.0553
PSOGPC Discontinuation	**39 from 123**	**91 from 220**	**14 from 78**	**5 from 55**
State-of-Art Discontinuation	11 from 123	25 from 220	7 from 78	3 from 55

working with a medical dataset where the negative class is prevalent, AUPRC can come in handy. A high AUC does not necessarily mean a high AUPRC, as it is possible for a model to have a high AUC but low AUPRC using the same data. The baseline AUPRC of a model is the percentage of positive values in the data (Table I). A curve is generated when the precision is plotted against the recall, and the area under the generated curve is AUPRC.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \tag{12}$$

III. RESULTS AND DISCUSSION

The examined datasets were randomly split into training and test sets on a ratio of 70% : 30% and the best regularization parameters for preventing overfitting were selected using a 5-fold cross validated grid search. Five standard classifiers were trained and their performance were compared with PSO-GPC. The classifiers are random forest (RF), categorical boosting (CB), LightGBM (LGB), XGBoost (XGB), and K-Nearest Neighbor (KNN). These classifiers performed differently despite some similarities which exist between them. The performance of PSO-GPC was further compared with those in published literature.

The performance analysis of the models on all examined cohorts is shown in Table II and the resulting curves in Fig. 3 (a)-(h). With the baseline AUPRC of Ascent-2, Celgene, Venice cohorts and All cohorts given as 0.2206, 0.0779, 0.0553, 0.1232 respectively, PSO-GPC provides better performance than other methods with considerable improvements. In view of the fact that the number of negative samples is greatly higher than the number of positive samples in the data, a high AUC does not necessarily mean a high AUPRC. When there is a high change in the number of false positives (i.e. patients who should continue with treatment but wrongly classified as patients who should discontinue treatment), it can only result in a low change in the FPR used for the ROC analysis. On the other hand, the precision reflects and showcase the effect which the high number of negative samples has on the performance of the classifier by making a comparison between the wrongly predicted positive (FP) and rightly predicted positives (TP) rather than the rightly predicted negative samples (TN) [39].

In articles published by [20], where a similar dataset was used, positive samples were about 10%. Submissions made for the challenge had the AUPRCs ranging between 0.088 and 0.178, and AUC ranging between 0.55 and 0.60 for All cohort. The AUPRCs are low due to the low number of predicted positive cases. In the submission made by [40] in a similar challenge (winners of the challenge), using random forest on All cohorts, Ascent-2, Celgene and Venice cohorts, the recorded AUCs were 0.6356, 0.5726, 0.6420, and 0.547 respectively, and AUPRCs were 0.2001, 0.3089, 0.1598, and 0.1006 respectively. The last two rows of Table III shows the estimated number of patients who can be prevented from wrong treatment. Using PSO-GPC, the number of patients to be discontinued from wrong treatment is triple (All and Ascent-2 cohort) and double (Celgene and Venice cohort) the number recorded by the state-of-art method. Analysis in the medical field aims towards consistency and best achievable performance. PSO-GPC has maintained consistency in performance across all analyzed cohorts with competitive AUCs and AUPRCs, performing better than results in published literature. We believe that the data preprocessing and optimized method used have influenced the improvement in our result compared with benchmark results published. Xgboost and LightGBM were recommended to perform better than random forest [40]. This recommendation influenced their use as classifiers for comparison. PSO-GPC performed consistently better across all examined datasets.

Kendall's coefficient of concordance used for the statistical ranking of the methods shows a high level of agreement in the ranking of the methods. Kendall's W ranges from 0 to 1, with 1 showing a high level of agreement between the raters. A score of 0.914, 0.971, 0.971, and 0.743 was obtained for All Cohort, Ascent-2, Celgene, and Venice cohort, respectively. PSO-GPC was consistently ranked with a mean average of 6.0 across the datasets. We further utilized the Friedman non-parametric test method in assessing/ranking the models for the datasets. The null hypothesis expects the average rank of the models to be equal. If they are not equal, the null hypothesis is rejected. With p-value $p < 0.001$, the mean rank obtained for the models is reported as PSO-GPC = 6.00, CB = 3.75, LGB = 4.13, RF = 2.63, KNN = 1.38 and XGB = 3.13. The consistent high rank of PSO-GPC shows its stability and robustness.

A. Feature Importance Analysis

We determined the important features in All cohort dataset using a random forest algorithm and AUPRC criterion. The

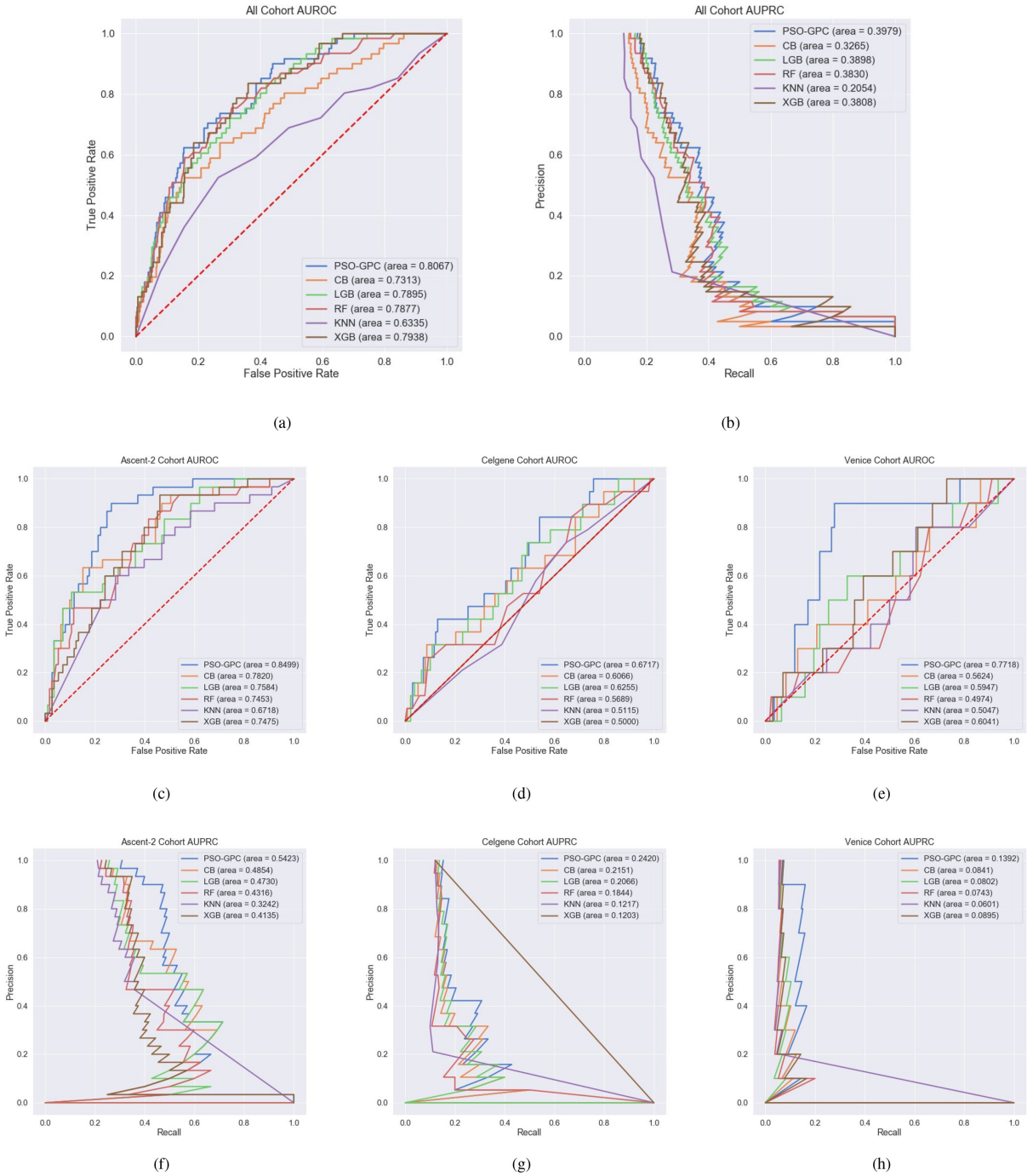

Fig. 3. Receiver Operating Characteristic (ROC) Curves with Area Under Curve (AUC) values of (a) All Cohort, (c) Ascent-2 Cohort, (d) Celgene Cohort, (e) Venice Cohort. Precision-Recall Curves with Area Under Curve (AUPRC) values of (b) All Cohort, (f) Ascent-2 Cohort, (g) Celgene Cohort, (h) Venice Cohort.

top-ranked 20 features are shown in Fig. 4 with their corresponding importance values. "ENDTRS _ C", the most important attribute has four categories. This attribute gives information about the reasons for discontinuation such as Adverse Effect, Possible Adverse Effect, Progression for patients still receiving treatment, and Complete for patients who completed the treatment. It essentially generates better information in the determination of patients who had to discontinue treatment. "ALB" represents Albumin which is a marker for assessing the nutritional status of patients. A low level of albumin can significantly affect the metastasis of malignant tumor cells [41]. Magnesium "MG" and Sodium "NA" are also important attributes that determine drug administration toxicity in patients. The deficiency of Sodium and Magnesium is known to cause an increase in risk for cancer patients [42], while a high concentration of Phosphorus "PHOS" on the other hand, can pose a potential risk for cancer [43].

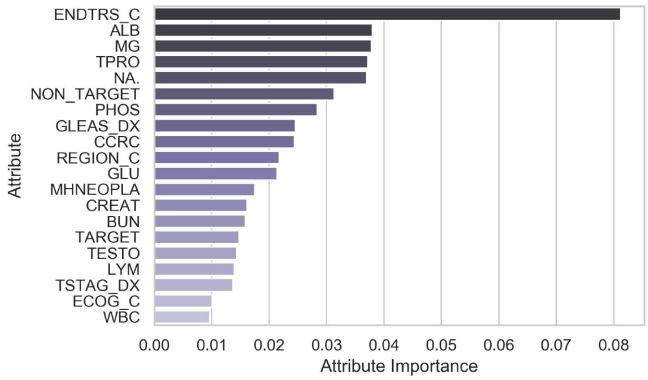

Fig. 4. Attribute Importance of All Cohort Dataset.

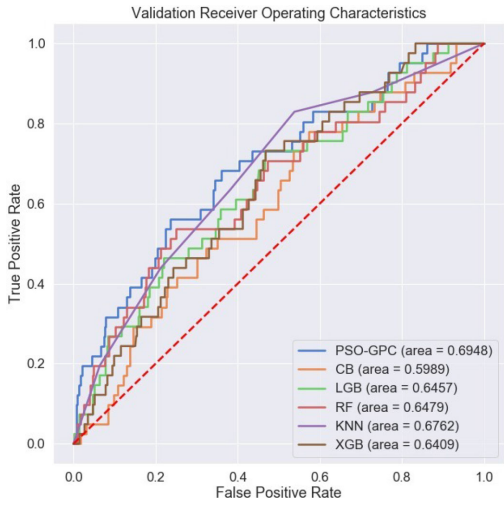

Fig. 5. Receiver Operating Characteristic (ROC) Curves on Independent Testing Dataset with Area Under Curve (AUC) values.

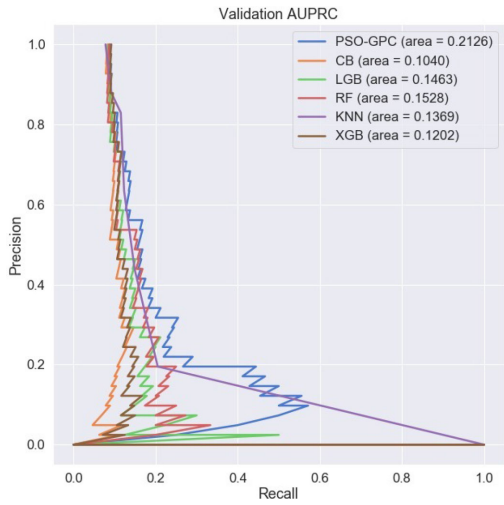

Fig. 6. Precision-Recall Curves of Independent Testing Dataset with Area Under Precision-Recall Curve (AUPRC) values.

Total Protein "TPRO" in the body can influence the growth of a tumor and toxicity of a drug, as it is evident that amino acid is one of the essential nutrients for the development of cancer cell [44], [45]. The non-target lesion, that is, the yet-to-be-measured noted changes in the tissue, are also essential factors for determining the degree of disease progression [46]. The Gleason score "GLEAS _ DX" is also a known strong determinant in the survival rate of prostate cancer patients. Creatinine clearance "CCRC" influences the survival rate of cancer patients [47]. A decrease in the creatinine clearance tends to increase the therapeutic effect by increasing the concentration of cisplatin. The environment/region "REGION _ C" where an individual resides influences the climatic condition, dietary pattern, and lifestyle of an individual. These factors can contribute to the body's reaction to treatment. Other important attributes in Fig. 4 are Glucose level, benign or malignant neoplasm, creatinine, blood urea nitrogen value, target lesion, testosterone level, lymphocytes value, primary tumor stage (i.e., score of tumor at primary location), patient's performance status, and white blood cells count. We note that majority of the top contributing features are related to the patient's laboratory medical history rather than the location of the tumor metastasis. We also consulted and discussed our findings with an expert to further affirm the importance of the stated features in the determination of patients reaction to treatment.

B. Independent Testing Set

Due to the low rate of discontinuation cases in the Celgene cohort dataset, we selected it as an independent testing data while we trained the models based on the combination of Ascent-2 and Venice cohort datasets. Attributes with importance values above the threshold of 0.001 were selected, and PSO-GPC was compared with other methods. An AUC of 0.6948 and AUPRC of 0.2126 were obtained, as shown in Fig. 5 and 6. This procedure of using only selected attributes was also implemented in the literature used for comparison. With the discontinuation rate of 7.79% in the Celgene cohort, which is equivalent to 0.0779 AUPRC baseline, if the usual therapy wrongly continues the treatment of 77.9 out of 1000 patient with docetaxel, it is expected that with PSO-GPC, we can discontinue the treatment of approximately 11 (i.e. $((0.2126 - 0.0779)/(1 - 0.0779) * 77.9)$) out of the approximately 78 patients, hence preventing them from adverse effects such as death. Considering that the state-of-art method could place approximately 10 out of 104 patients from wrong treatment, PSO-GPC performs better, considering that it can discontinue wrong treatment of 15 out of 104 patients. Despite training with lesser data, PSO-GPC harnesses its strategy of updating the particles with the global best solution for global search. This strategy enables the optimization of the Gaussian process classifier in selecting suitable parameters for robust performance.

C. Contrast With Related Works

In the challenges where the same data had been used, the majority of submissions from teams were based on ensembles and standard classifiers. In previous studies, ensembles have shown

superior predictive ability compared to standard classifiers. We believe that random forest, which is an ensemble, contributed to the overall performance of [40] in the challenge. Methods used by [20], also involved standard classifiers and ensembles. In our study, using a standard classifier that is optimized with the metaheuristic algorithm enables the white-box robustness of the method to different data conditions. To achieve the robust performance, harnessing the power of metaheuristic algorithms in the optimization of standard classifiers is a viable prospect.

IV. Conclusion

The prevalence of prostate cancer in men and its high mortality rate necessitates a pragmatic approach to it. Treatment with docetaxel is prevalent in patients with mCPRC, but it sometimes has an adverse effect on some patients due to toxicity. Based on the design of PSO-GPC, we have successfully predicted the discontinuation of docetaxel treatment.This method has performed consistently better than examined classifiers (including XGB and LGB) and methods in the literature. With the 1600 patient records, its unbiased and robust performance has been demonstrated under diverse experimental settings. We have also revealed the contribution of clinical values in patients' records in the determination of treatment discontinuation. The limitation lies in the large number of missing values whose records cannot be removed to prevent bias [48], but had to be imputed using suitable data imputation techniques. Sufficient data size is important for cancer research. This helps the model to train with diverse samples and mitigate any effect that might arise from missing data. We believe such challenges of missing data can be addressed right from the data capturing and documentation. This will produce research outcomes that have high clinical importance. In future works, new techniques can be developed for imputation. Adequate data capturing and documentation should also be prioritized.

Acknowledgment

The authors are grateful to the contributors of the patient data and Project Data Sphere for granted access. The authors appreciate Dr. Adebimpe Yusuf for the consultation.

References

[1] L. Rahib, B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, and L. M. Matrisian, "Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States," *Cancer Res.*, vol. 74, no. 11, pp. 2913–2921, 2014.

[2] R. A. Castillejos-Molina and F. B. Gabilondo-Navarro, "Cácer of próstata," *public health of méxico*, vol. 58, no. 2, pp. 279–284, 2016.

[3] A. W. Lambert, D. R. Pattabiraman, and R. A. Weinberg, "Emerging biological principles of metastasis," *Cell*, vol. 168, no. 4, pp. 670–691, 2017.

[4] E. B. Rankin and A. J. Giaccia, "Hypoxic control of metastasis," *Science*, vol. 352, no. 6282, pp. 175–180, 2016.

[5] S. Lee, S. Kerns, H. Ostrer, B. Rosenstein, J. O. Deasy, and J. H. Oh, "Machine learning on a genome-wide association study to predict late genitourinary toxicity after prostate radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 101, no. 1, pp. 128–135, 2018.

[6] B. Qian, Y. Yao, C. Liu, J. Zhang, H. Chen, and H. Li, "SU6668 modulates prostate cancer progression by downregulating MTDH/AKT signaling pathway," *Int. J. Oncol.*, vol. 50, no. 5, pp. 1601–1611, 2017.

[7] E. D. Crawford *et al.*, "The role of therapeutic layering in optimizing treatment for patients with castration-resistant prostate cancer (prostate cancer radiographic assessments for detection of advanced recurrence II)," *Urology*, vol. 104, pp. 150–159, 2017.

[8] M. Tucci, G. V. Scagliotti, and F. Vignani, "Metastatic castration-resistant prostate cancer: Time for innovation," *Future Oncol.*, vol. 11, no. 1, pp. 91–106, 2015.

[9] F. Saad and S. J. Hotte, "Guidelines for the management of castrate-resistant prostate cancer," *Can. Urological Assoc. J.*, vol. 4, no. 6, pp. 380–384, 2010.

[10] M. S. Litwin and H.-J. Tan, "The diagnosis and treatment of prostate cancer: A review," *JAMA*, vol. 317, no. 24, pp. 2532–2542, 2017.

[11] T. M. Beer *et al.*, "Enzalutamide in men with chemotherapy-naive metastatic castration-resistant prostate cancer: Extended analysis of the phase 3 prevail study," *Eur. Urol.*, vol. 71, no. 2, pp. 151–154, 2017.

[12] S. Niraula *et al.*, "Duration of suppression of bone turnover following treatment with zoledronic acid in men with metastatic castration-resistant prostate cancer," *Future Sci. OA*, vol. 4, no. 1, 2017, Art. no. FSO253.

[13] J. Mateo *et al.*, "Olaparib in patients with metastatic castration-resistant prostate cancer with DNA repair gene aberrations (TOPARP-B): A multicentre, open-label, randomised, phase 2 trial," *Lancet Oncol.*, vol. 21, no. 1, pp. 162–174, 2020.

[14] P. Nuhn *et al.*, "Update on systemic prostate cancer therapies: Management of metastatic castration-resistant prostate cancer in the era of precision oncology," *Eur. Urol.*, vol. 75, no. 1, pp. 88–99, 2019.

[15] A. Templeton *et al.*, "Translating clinical trials to clinical practice: Outcomes of men with metastatic castration resistant prostate cancer treated with docetaxel and prednisone in and out of clinical trials," *Ann. Oncol.*, vol. 24, no. 12, pp. 2972–2977, 2013.

[16] P. Cornford *et al.*, "EAU-ESTRO-SIOG guidelines on prostate cancer. Part II: Treatment of relapsing, metastatic, and castration-resistant prostate cancer," *Eur. Urol.*, vol. 71, no. 4, pp. 630–642, 2017.

[17] R. Cuocolo *et al.*, "Machine learning applications in prostate cancer magnetic resonance imaging," *Eur. Radiol. Exp.*, vol. 3, no. 1, pp. 1–8, 2019.

[18] M. Antonelli *et al.*, "Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists," *Eur. Radiol.*, vol. 29, no. 9, pp. 4754–4764, 2019.

[19] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean, "Deep learning-based Gleason grading of prostate cancer from histopathology images–Role of multiscale decision aggregation and data augmentation," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 5, pp. 1413–1426, May 2020.

[20] F. Seyednasrollah *et al.*, "A dream challenge to build prediction models for short-term discontinuation of docetaxel in metastatic castration-resistant prostate cancer," *JCO Clin. Cancer Informat.*, vol. 1, pp. 1–15, 2017.

[21] H. Guo, U. Kruger, G. Wang, M. K. Kalra, and P. Yan, "Knowledge-based analysis for mortality prediction from CT images," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 457–464, Feb. 2020.

[22] K. Praveena, K. Bhargavi, and K. Yogeshwari, "Comparision of PSO Algorithm and Genetic Algorithm in WSN using NS-2," in *Proc. Int. Conf. Curr. Trends Comput., Elect., Electron. Commun.*, IEEE, 2017, pp. 513–516.

[23] X.-N. Ma and H. Wang, "Topic trends forecasting using BP neural network model based on PSO," *Comput. Eng. Des.*, p. 09, 2018.

[24] "Project data sphere," Accessed: Mar. 4, 2020. [Online]. Available: https://www.projectdatasphere.org/

[25] F. Santore, E. C. d. Almeida, W. H. Bonat, E. H. Pena, and L. E. S. de Oliveira, "A framework for analyzing the impact of missing data in predictive models," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 2209–2212.

[26] S. Armijo-Olivo, W. Machalicek, L. Dennett, and N. Ballenberger, "Influence of attrition, missing data, compliance, and related biases and analyses strategies on treatment effects in randomized controlled trials in rehabilitation: A methodological review," *Eur. J. Phys. Rehabil. Med.*, vol. 56, no. 6, pp. 799–816, 2020.

[27] I. Cornelisz, P. Cuijpers, T. Donker, and C. van Klaveren, "Addressing missing data in randomized clinical trials: A causal inference perspective," *PLoS One*, vol. 15, no. 7, 2020, Art. no. e0234349.

[28] Z. Xu *et al.*, "Software defect prediction based on Kernel PCA and weighted extreme learning machine," *Inf. Softw. Technol.*, vol. 106, pp. 182–200, 2019.

[29] T. H. Nguyen *et al.*, "Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning," *J. Biomed. Opt.*, vol. 22, no. 3, 2017, Art. no. 036015.

[30] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci.*, vol. 513, pp. 429–441, 2020.
[31] C. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2. Cambridge, MA, USA: MIT Press, vol. 493, 2006, p. 494.
[32] M. Kuss and C. E. Rasmussen, "Assessing approximate inference for binary Gaussian process classification," *J. Mach. Learn. Res.*, vol. 6, no. 10, pp. 1679–1704, 2005.
[33] J. Riihimäki *et al.*, "Laplace approximation for logistic Gaussian process density estimation and regression," *Bayesian Anal.*, vol. 9, no. 2, pp. 425–448, 2014.
[34] Y.-J. Gong *et al.*, "Genetic learning particle swarm optimization," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2277–2290, Oct. 2016.
[35] H. Iiduka, "Stochastic fixed point optimization algorithm for classifier ensemble," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4370–4380, Oct. 2020.
[36] K.-L. Du and M. Swamy, "Particle swarm optimization," in *Search and Optimization by Metaheuristics*, Cham, Switzerland: Springer, 2016, pp. 153–173.
[37] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: classification evaluation," *Nature Methods*, vol. 13, no. 8, pp. 603–604, 2016.
[38] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, vol. 10, no. 3, 2015, Art. no. e0118432.
[39] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.
[40] K. Deng, H. Li, and Y. Guan, "Treatment stratification of patients with metastatic castration-resistant prostate cancer by machine learning," *Iscience*, vol. 23, no. 2, 2020, Art. no. 100804.
[41] J. Ying, D. Zhou, T. Gu, J. Huang, and H. Liu, "Pretreatment albumin/fibrinogen ratio as a promising predictor for the survival of advanced non small-cell lung cancer patients undergoing first-line platinum-based chemotherapy," *BMC Cancer*, vol. 19, no. 1, pp. 1–8, 2019.
[42] B. Oronsky *et al.*, "Electrolyte disorders with platinum-based chemotherapy: Mechanisms, manifestations and management," *Cancer Chemotherapy Pharmacol.*, vol. 80, no. 5, pp. 895–907, 2017.
[43] J. J. Anderson, "Potential health concerns of dietary phosphorus: Cancer, obesity, and hypertension," *Ann. New York Acad. Sci. USA*, vol. 1301, no. 1, pp. 1–8, 2013.
[44] S. M. Davidson *et al.*, "Direct evidence for cancer-cell-autonomous extracellular protein catabolism in pancreatic tumors," *Nature Med.*, vol. 23, no. 2, pp. 235–241, 2017.
[45] X. Huang, H. Zhang, X. Guo, Z. Zhu, H. Cai, and X. Kong, "Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) in cancer," *J. Hematol. Oncol.*, vol. 11, no. 1, pp. 1–15, 2018.
[46] S. Raskin, E. Klang, and M. Amitai, "Target versus non-target lesions in determining disease progression: Analysis of 545 patients," *Cancer Imag.*, vol. 15, pp. 1–1, 2015, Art. no. S1:S8.
[47] S. Gözel *et al.*, "P1. 07-010 influence of creatinine clearance on survival parameters in small cell lung cancer treated with Cisplatin-based chemotherapy regimen: Topic: Drug treatment alone and in combination with radiotherapy," *J. Thoracic Oncol.*, vol. 12, no. 1, pp. S701–S702, 2017.
[48] D. X. Yang *et al.*, "Prevalence of missing data in the national cancer database and association with overall survival," *JAMA Netw. Open*, vol. 4, no. 3, p. e211793, 2021.