Neural Computing and Applications (2024) 36:1585–1597
https://doi.org/10.1007/s00521-023-09142-3

ORIGINAL ARTICLE

Bladder cancer gene expression prediction with explainable algorithms

Kevser Kübra Kırboğa[1,2]

Received: 26 May 2023 / Accepted: 16 October 2023 / Published online: 11 November 2023

Abstract
In this study, we aimed to classify bladder cancer patients using tumoral and non-tumoral gene expression data. In this way, we aimed to determine which genes are effective on tumoral and normal tissues. In addition, for this purpose, we planned to perform this classification using interpretable methods (The aim of this study was to classify bladder cancer patients using gene expression data from tumoral and non-tumoral tissues. By doing so, we wanted to determine which genes were effective on both tumoral and normal tissues. Moreover, for this purpose, we planned to use interpretable methods for this classification.). Analyses using permutation feature importance (PFI), SHapley Additive exPlanations (SHAP), local interpretable model-agnostic explanations (LIME), and Anchor methods on data from Gene Expression Omnibus (GEO) and Curated Microarray Database we did (We performed analyses using permutation feature importance (PFI), SHapley Additive exPlanations (SHAP), local interpretable model-agnostic explanations (LIME), and Anchor methods on data from Gene Expression Omnibus (GEO) and Curated Microarray Database.). These are eXplainable methods used to determine the importance of genes in classification. According to the results of our study, the most important genes were determined as LINC00161, ACACB, and CBARP according to the PFI method, HSPA6, STON2, and RFC2 according to the SHAP method, PRUNE2 and ABCC13 according to the LIME method, and TMEM74, KLHL10, and GAMT according to the Anchor method. This study shows that genes involved in other cancer types are also effective in bladder cancer. In addition, it has been observed that using explainable methods in cancer data can support prognosis and treatment in the clinic.

Keywords Bladder cancer · Gene expression · Explainable artificial intelligence · Shapley explanations · Anchor · Lime

1 Background

According to the World Health Organization (WHO), over 200,000 people die yearly from bladder cancer, affecting around 600,000 individuals globally [1]. Bladder cancer is a type of cancer that can be treated with early detection and treatment. However, it can be fatal if diagnosed in advanced stages and untreated. For this reason, the death rate of bladder cancer cases may vary depending on the stage and treatment of the cancer. Genetic and epigenetic factors play an essential role in the development and progression of bladder cancer. These factors can cause changes in tumour suppressor genes or oncogenes, disruption of DNA repair mechanisms, altered levels of DNA methylation, and differentiation of gene expression patterns. Gene expression data are molecular measurements that reflect the process from the DNA sequence of a gene to the production of RNA and protein [2]. Analysis of gene expression data is an essential tool for revealing the functions, interactions and associations of genes with diseases [3, 4]. Therefore, a gene expression study is critical to identify molecular subtypes of bladder cancer and prognostic and therapeutic targets. Sozeri et al. analysed gene expression data of primary bladder cancer and bladder cancer cell lines with ELF3 mutations. Their analysis showed that the

✉ Kevser Kübra Kırboğa
kubra.kirboga@bilecik.edu.tr

1 Bilecik Seyh Edebali University, Faculty of Engineering, Bioengineering Department, Bilecik 11100, Türkiye

2 Informatics Institute, Istanbul Technical University, Maslak, Istanbul 34469, Türkiye

ELF3 mutation disrupted genes involved in cell migration and cell–cell junction organization [5]. Zaravinos et al., on the other hand, analysed samples taken from 30 human bladder cancer and adjacent normal tissue with microarray technology in bladder cancer.

In particular, the authors drew attention to the expression changes due to the multiple biological functions of a molecule called osteopontin (OPN) [6]. Artificial intelligence (AI) and machine learning (ML) methods are widely used to analyze gene expression data [3, 4, 7]. These methods can process, classify, cluster, filter, and visualize data [7–9]. AI and ML methods can also identify biomarkers and drug targets for disease diagnosis, prognosis and treatment based on gene expression data. Chen et al. used integrated bioinformatics and experimental testing to identify and validate central genes associated with bladder cancer. By analysing two mRNA microarray datasets, the authors identified 362 genes and 13 significant genes with differential expression in bladder cancer. They suggested these genes could be potential biomarkers for diagnosing and treating bladder cancer [10]. Wagner showed that it is possible to predict gene expression from DNA sequences using a neural network, an AI model, in 2022 for gene expression and AI studies.

Along with high-volume experiments, Wagner trained the neural network and predicted the activity and evolution of gene promoter sequences [11]. Abbod et al. collected gene expression data from bladder cancer patients and healthy individuals and processed these data with AI and ML algorithms. The authors used classification methods such as support vector machines (SVM), artificial neural networks (ANN), and decision trees (DT) to distinguish between bladder cancer patients and healthy individuals. The results revealed that the SVM method outperformed other methods. They also used gene selection methods to identify genes that play an essential role in bladder cancer. This study was a computational study that could contribute to diagnosing and treating bladder cancer [12].

Although the importance of gene expression data on bladder cancer has been studied with various AI and ML methods, the use of model-agnostic methods on gene data is insufficient. Model agnostic methods are used to increase ML models' interpretability. These methods measure the importance or impact of features to explain a model's predictions. Model agnostic methods can be applied to any model, regardless of the model's structure or algorithm. PFI, SHAP, LIME, and Anchor are model-agnostic methods used to interpret ML models. These methods measure the importance or influence of features in the predictions of a model. Model-agnostic methods can be applied to any model, regardless of the model's structure or algorithm. These methods benefit gene expression data analysis because they can help identify genes associated with disease or phenotype. PFI, permutation feature importance, is a method that measures the decrease in the model's performance when a feature is randomly mixed. The rationale is that if a feature is important to the prediction, mixing its values will worsen it. PFI can rank features according to their importance and select the most relevant for further analysis [13]. This method has been used in the effects of heterologous different COVID-19 vaccines on immune cells [14], using miRNA profiling as a method in the diagnosis of Meniere's disease [15] and genome-wide examination of gene-immune relationships to predict new therapeutic targets in oncology [16]. SHAP, Shapley additive explanations, is a method that assigns a Shapley value to each attribute that contributes to a prediction [17, 18]. The Shapley value is calculated based on cooperative game theory, which considers all possible combinations of features and their marginal effects. SHAP can be used to explain the predictions of a model and identify features that have a positive or negative effect on the outcome. In 2023, Sanchez et al. used the XAI and SHAP methods to identify biomarkers for squamous cell carcinoma (SCC), one of the skin cancer types. As a result of the study, it was shown that the XAI model can distinguish between SCC patients and healthy individuals with 96% accuracy and identifies 15 SCC-associated miRNAs. Some of these miRNAs have been previously associated with SCC in the literature, while others have been newly discovered. This study reveals that XAI can be useful in identifying biomarkers for SCC [19]. In another study proving that XAI can be a useful tool in the identification of biomarkers for breast cancer, Kumar et al., in the detection of biomarkers derived from peripheral blood mononuclear cells (PBMC), XAI, which consists of two components, such as variational autoencoder (VAE) and decision tree (DT). They created the SHAP model [20]. The SHAP method generated a new prognostic signature for AML patients based on ferroptosis-related genes [21]. Another model-agnostic method used in this study is LIME (local interpretable model-agnostic explanations). This method provides a locally linear approximation of a model around a given forecast [22]. The logic is that a complex model can be simplified in a small neighbourhood of the input space and explained by a linear model. LIME can be used to describe a model's predictions and identify features with high weights in the local linear model. Lai and his team developed six ML algorithms to predict the course of AD and identify distinctive genes while aiming to identify immune microenvironment subtypes and distinctive genes in the diagnosis and risk prediction of Alzheimer's disease (AD). The output of ML models was interpreted using SHAP and LIME algorithms [23]. Oni and Qiao, which provides a useful framework to increase the interpretability of ML models and provide transparency in the biomedical field, have developed an ML model that can distinguish between eight cancer types using RNA-seq expression and single nucleotide variation (SNV) data [24]. Similar to our study, in a study that applied ML methods using DNA methylation data to classify 24 cancer types using public databases such as TCGA and GEO, distinctive genes were determined for cancer types with the LIME method [25]. Anchor explanations we used in our study, on the other hand, are a method that finds the least number of features with high confidence in predicting the result [26]. The logic of the method is that an Anchor is a rule that is always true for a given prediction and is valid regardless of other properties. Anchor can be used to explain the predictions of a model and identify features that are sufficient and necessary for the result. Model agnostic methods are an important tool to explain the decisions of ML models to classify cancer types using gene expression data. These methods allow the model to classify which genes are expressed, and to reveal the biological meanings of these genes. However, there are very few studies in the literature using model agnostic methods in gene expression studies. Specifically, there are no such studies for bladder cancer. Therefore, our study is unique and distinctive to classify bladder cancer types and identify discriminating genes using LIME and Anchor annotations. The main research question of this study is how bladder cancer types can be classified based on gene expression data and what are the molecular mechanisms behind this classification. The hypothesis is that a classification model developed with AI and ML methods can distinguish bladder cancer types with high accuracy and the decisions of the model interpreted with model agnostic methods can reveal bladder cancer-related genes, their biological meanings and predictive values. The study aims to classify bladder cancer types based on gene expression data and to explain the molecular mechanisms behind this classification. For this purpose, we developed a classification model using AI and ML methods and interpreted the decisions of our model using model agnostic methods PFI, SHAP, LIME and Anchor. Also, in this study, we aimed to determine the genes associated with bladder cancer, their biological significance and prognostic value. This study is the first to use model agnostic methods for bladder cancer and offers a new way to better understand the molecular mechanisms and biomarkers of bladder cancer. This study makes an important and valuable contribution to improving the diagnosis, prognosis and treatment options of bladder cancer patients. In this study, a dataset (GSE31189) containing gene expression data from bladder cancer patients in the GEO database was used. This dataset shows that the study is consistent with its purpose and scope. Because, with this data set, it is possible to classify bladder cancer types and explain the molecular mechanisms behind this classification.

In the second part of the article, we discuss the theoretical foundations of artificial intelligence (AI) and machine learning (ML) methods we use to classify bladder cancer types and the model agnostic methods we use to explain the decisions of our classification model and identify bladder cancer-related genes, such as PFI, SHAP, LIME and Anchor methods. We detail the GEO dataset's source, characteristics, and preprocessing steps, including gene expression data from bladder cancer patients based on working principles. In the third part of the article, we evaluate the performance of our classification model with various metrics and present the explanations obtained by model-agnostic methods visually and numerically. This section shows the distinctive genes for bladder cancer types and their severity. In the fourth part of the article, we discuss our findings' biological significance and prognostic value by comparing them with the literature. This section analyses bladder cancer-related genes' functions, interactions, and roles in disease mechanisms. In the fifth part of the article, we summarize our study's contributions, limitations, and suggestions for future studies. In this section emphasises that our research demonstrates the potential value of AI and ML methods in diagnosing and treating bladder cancer.

2 Experiments

The research consists of data collection and preprocessing, model development, and implementation of ML and XAI. Python version 3.11 was employed for conducting the research. The libraries used throughout the performance were Matplotlib version 3.7.1, pip version 22.0.4, Sklearn version 1.2.2, Pandas version 2.0.1, RDkit version 2023.3.1, Shap version 0.41.0, eli5 version 0.13.0, sci-kit-plot version 0.3.7, and NumPy version 1.24.3. The implementation was carried out on a computer with Intel® Core ™ i5-8300H CPU 2.30GHz, 64-bit operating system, × 64-based processor, and 32 GB RAM.

2.1 Data preprocessing

The bladder cancer dataset includes 92(52-Cancer Urothelia) microarrays for Homo sapiens, selected by carefully reviewing over 30,000 microarray experiments using rigorous filtering criteria and downloaded from the GEO(GSE31189) [27] and Curated Microarray Database (CuMiDa) [28] databases (Fig. 1). The dataset was manually edited for background correction, normalisation, sample quality analysis, and removal of faulty probes [28]. In addition, the Affymetrix symbols in the dataset have been translated into gene names using The Database for Annotation, Visualisation, and Integrated Discovery

(DAVID), a comprehensive set of functional annotation tools for researchers to understand the biological meaning behind extensive gene lists [29].

2.2 Model development

This study used different ML algorithms for a tumoral or non-tumoral prediction from bladder cancer gene expression data. The algorithms used are RandomForestClassifier, XGBoost, DecisionTreeClassifier, KNeighborsClassifier, SVC, GaussianNB, and LogisticRegression. In addition, different metrics such as Train Accuracy, Test Accuracy, Precision, Recall, F1-score, Brier Score, and AUC were used to evaluate the accuracy of the models. Training Accuracy measures how well the model fits the training data (Table 1). It is the percentage of correctly classified samples in the training set. Test Accuracy counts how well the model can generalise to new data. The rate of samples correctly classified in the test set. Precision measures the percentage of true positives among all predicted positives. Recall measures the percentage of predicted positives among all true positives. The F1-score calculates the harmonic mean of precision and recall. The Brier Score measures the mean squared difference between the predicted probabilities and the actual results. The AUC measures the area under the receiver operating characteristic curve and plots the true positive rate against the false positive rate at various threshold settings. These metrics can help us compare different models and choose the best fit for our problem. It also helps us understand what aspects of the model's performance need improvement and how to improve it [30–32]. The model development process included typical ML steps such as splitting the dataset, tuning the model hyperparameters, and selecting the model.

RandomForestClassifier: This ML algorithm creates multiple decision trees and makes these trees predict by majority vote. It has advantages such as reducing overfitting decision trees, being resistant to noise in data and providing high accuracy. However, this algorithm may use a lot of memory, have a long training time, and require parameter tuning [33]. In gene expression data, the RandomForestClassifier algorithm has been used in applications such as breast cancer diagnosis [34] and the classification of many cancer types [35–37].

XGBoost: It is an ML algorithm known as *extreme gradient boost*. This algorithm can solve classification and regression problems using a collection of decision trees. It has advantages, such as fast training time, high performance, scalability and regularisation. However, this algorithm contains too many parameters, can be complex, and may be prone to overfitting. [38–40].

DecisionTreeClassifier: This algorithm, which works by dividing the data with simple decision rules, has the advantages of being easily interpreted, tolerant of missing values, processing categorical values and capturing interactive effects. However, this algorithm can be sensitive to overfitting, unstable, and high variance [41].

KNeighborsClassifier [42]**:** It is a simple, flexible and nonparametric method that allows an object to be classified according to the labels of its nearest neighbours. It can also model complex boundaries and be applied to multiple classification problems. However, this algorithm does not perform well on high-dimensional data, spends too much time calculating distances, and can be affected by unbalanced data distributions [42–44].

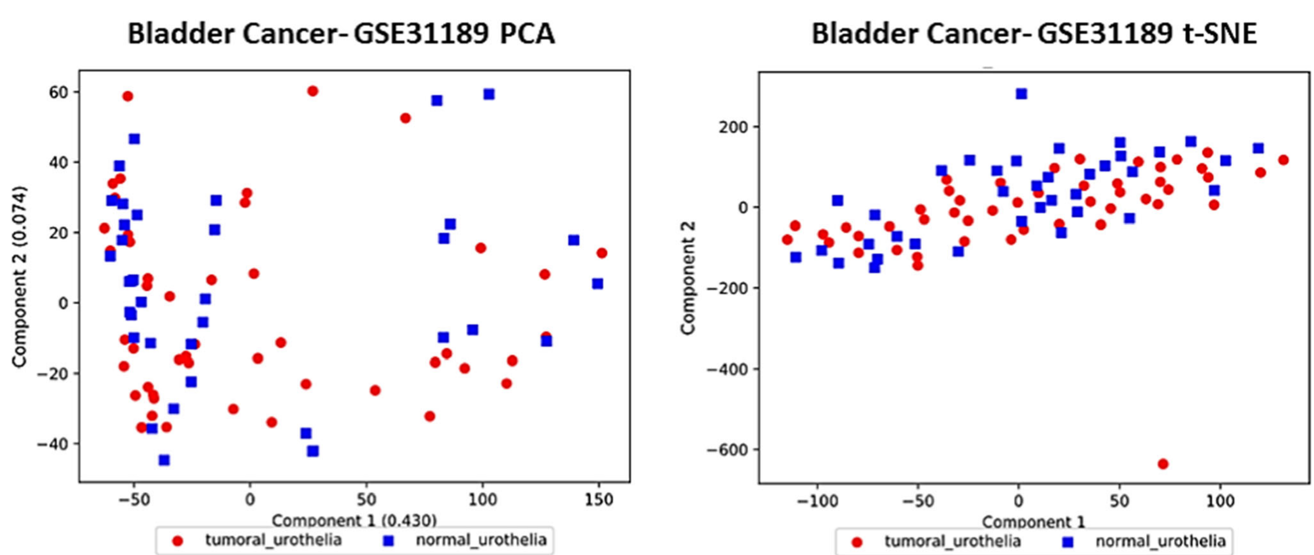

Fig. 1 PCA (Principal Component Analysis) charts were created to show the overall distribution or correlation of the data set used in the study, and t-SNE (t-distribution Stochastic Neighbor Embedding) plots to show clustering trends in the data

SVC: Support vector machines try to find a separation hyperplane to separate data into two or more classes. It has advantages like high accuracy, working well with small training sets, and modelling nonlinear bounds using kernel functions. It contains too many parameters, can take long training time, can be complex in multi-classification problems, and may not perform well on unbalanced datasets [45–47].

GaussianNB: GaussianNB algorithm, used for continuous features following the Gaussian distribution, is a simple, fast and probabilistic method. It can also be applied to multiple classification problems and requires little training data. However, this algorithm assumes features are independent, do not perform well on high-dimensional data, and can be affected by unstable datasets [48, 49].

LogisticRegression: LogisticRegression is a simple, interpretable and regularise method for estimating the probability that an object belongs to two or more classes. It can also be applied to multiple classification problems and model nonlinear boundaries using kernel functions. However, this algorithm assumes that the features are linearly separable, can be sensitive to overfitting and will not perform well on unbalanced datasets [50].

2.3 Permutation feature importance

Permutation feature importance (PFI) is a method used to measure the importance of variables (features) in ML models [13, 51]. This method evaluates each feature's impact on any model's performance. The PFI method works by redrawing random samples from a feature and measuring the change in model accuracy. The severity of a gene depends on the reduction in model accuracy by removing that gene with random redraws. If there is a significant decrease in model accuracy when genes are removed, this gene has an essential role in predicting the model correctly. This method suits bladder cancer prediction because the model must identify the most influential genes. This method evaluates in detail the impact of each gene in the dataset on prediction performance. According to their results, the 20 most significant genes are the variables that contribute the most to the model's best performance. Therefore, these genes were chosen to increase the model's accuracy or help explain the model more intelligibly.

2.4 Explainable AI

XAI is the ability of an AI model to clearly explain the causes and effects of input variables in the decision-making process. This means that the computational approaches in the model are understandable, cause-effect relationships in data processing can be traced, and it can be determined on which features the model makes a decision. XAI models are significant for vital decisions such as medical diagnostic evaluations and crime prevention. SHAP (SHApley Additive exPlanations) is a feature materiality method. SHAP is inspired by a game theory known as Shapley values [18]. Shapley values determine a player's contribution to a coalition's earnings. SHAP measures feature importance by combining feature weights with Shapley values. SHAP helps understand the relationship between a feature and a prediction. LIME (Local Interpretable Model-agnostic Explanations) is one of the XAI methods and explains the reasons for the predictions of a model for a given sample [52]. LIME creates a local model to understand the predictions of a model and uses that local model to predict its properties. This method explains complex models that work like a magic box. Anchor creates simple, independent, human-understandable rules that define certain conditions for a model's predictions [53, 54]. These rules can be used to understand the reason for the prediction. The Anchor method helps to identify the features required for the predictions of a model and also reveals the reasons that explain the model's decision. In this study, since a tree-based algorithm was chosen for SHAP analysis, are used *TreeShap*, *lime.lime_tabular* for LIME, *anchor-exp* for Anchor, and *permutationimportance* libraries for PFI were used.

3 Results

3.1 Model development

Various evaluation metrics were used to evaluate multiple model algorithms within the scope of the study. These metrics include accuracy of predictions made on training and test data, precision, recall, F1-score, Brier score, and Area Under the Curve (AUC). Based on the results obtained, it has been observed that the best performance is exhibited by the RandomForestClassifier algorithm(as shown Table 1). The high fit shown to the training data and the high results of the prediction success in the test data reveal its superiority compared to other algorithms. Therefore, as a result of the evaluation metrics analysis, it has been determined that the most appropriate choice of model selection is the RandomForestClassifier algorithm.

The AUC values calculated from the ROC curves analysis show that the RandomForestClassifier (0.67) algorithm provides the best classification performance. In addition, these values indicate that the model has a high rate of correct classification and performs better than other models (Fig. 2).

3.2 Permutation feature importance

As a result of the analysis made with the permutation feature importance (PFI) method, the essential genes in cancer prediction and the importance values of these genes were determined. The significance of genes indicates the magnitude of their predictive effects. Therefore, genes with higher significance values play a more significant role in cancer prediction. According to the data obtained from the analysis, the 20 most significant genes were determined (Figure 3) (Table 2). These results can help make more accurate and reliable predictions by selecting which genes are more effective in predicting cancer.

This list contains the most important genes for predicting bladder cancer and the weights assigned to these genes. The importance of each gene's contribution to inference was calculated using the PFI method. The essential gene was identified as LINC00161 and had the highest weight value (0.06%). Other essential genes include ACACB, PRSS33, MAPK1, FRAS1, CLEC12A, and GAMT. The importance of these genes is thought to play a decisive role in the diagnosis or prognosis of bladder cancer (Fig. 3) (Table 2).

Heatmap, a method used to visualise gene expression levels in the bladder cancer gene expression dataset, is the visualisation of gene expression data in the form of a heat map. Gene expressions are represented by assigning gene expression levels to an increasing or decreasing temperature colour scale. Therefore, visual similarities and differences between gene expressions can be easily noticed. In this study, heatmap maps of the 20 most essential genes obtained using the PFI method were drawn (Fig. 4). This visualisation allows the expression levels of each gene in the dataset to be represented in different temperature colours, providing a quick understanding of the visual similarities and differences between these genes. For example, 20 genes are listed on the left, with each observation (data point) listed above. The colour scale measures the standard deviations of genes over observations. Darker regions indicate that genes with higher standard deviations have a more substantial effect. Therefore, this heatmap visualises the PFI results, showing in which observations the most critical genes had the most significant impact.

3.3 Explainable AI

SHAP values are given on the horizontal axis of Fig. 5, showing the effect each gene has on the output. Positive SHAP values indicate features that increase model output, while negative SHAP values indicate features that decrease model output. Each bar of the graph is a gene, and the gene's SHAP value determines the length of the bar. According to ShAP results, it was seen that the gene with the highest positive SHAP value was "HSPA6". This means this gene is the trait that increases the model output the most. Other essential genes are "STON2", "RFC2," and "PRSS33," respectively. This information can help us determine which genes significantly influence model output and may be useful in selecting these genes as targets for cancer therapy.

Figure 6 shows the effect of different genes in the bladder cancer dataset on the prediction of bladder cancer. The orange indicates that the value of that gene has a high positive impact, while the dark blue indicates that the value of that gene has a high adverse effect. While it has been seen that genes such as PRUNE2, ABCC13, and CLEC12A in blue colour may cause tumour bladder cancer, genes such as COP1, GARIN4, and FRAS1 have been found to affect the non-tumoral urethra.

3.4 Anchor explainability

This study applied the "Anchor" model description to bladder cancer gene expression data. This description specifies the conditions for a particular sample on this run. In this case, the conditions are: "TMEM74" must be greater than "2.36", "KLHL10.1" must be less than "3.61", and "GAMT" must be greater than "5.51". In an

Table 1 Metric value results of developed ML algorithms

Model algorithms	Train accuracy	Test accuracy	precision	recall	f1-score	Brier score	AUC
RandomForestClassifier	1.0	0.56	0.53	0.53	0.52	0.46	0.67
XGBoost	1.0	0.55	0.56	0.56	0.53	0.46	0.64
DecisionTreeClassifier	1.0	0.53	0.52	0.53	0.53	0.33	0.64
KNeighborsClassifier	0.76	0.53	0.53	0.53	0.52	0.46	0.53
SVC	0.71	0.46	0.54	0.53	0.44	0.53	0.55
GaussianNB	0.8	0.53	0.53	0.53	0.52	0.46	0.59
LogisticRegression	0.85	0.4	0.42	0.42	0.40	0.6	0.44

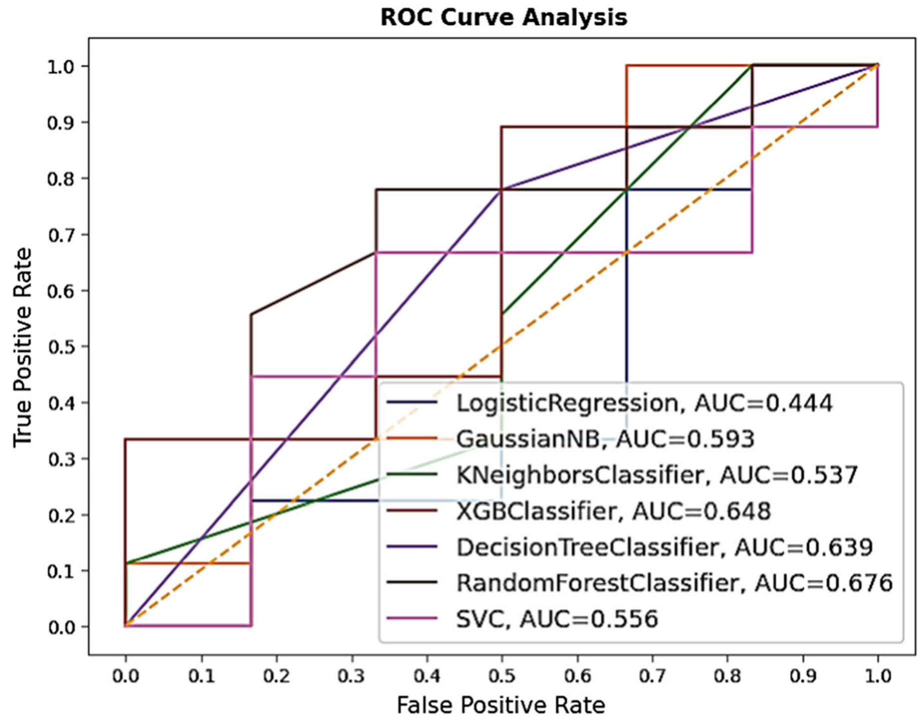

Fig. 2 AUC values and ROC curves of the seven ML models developed

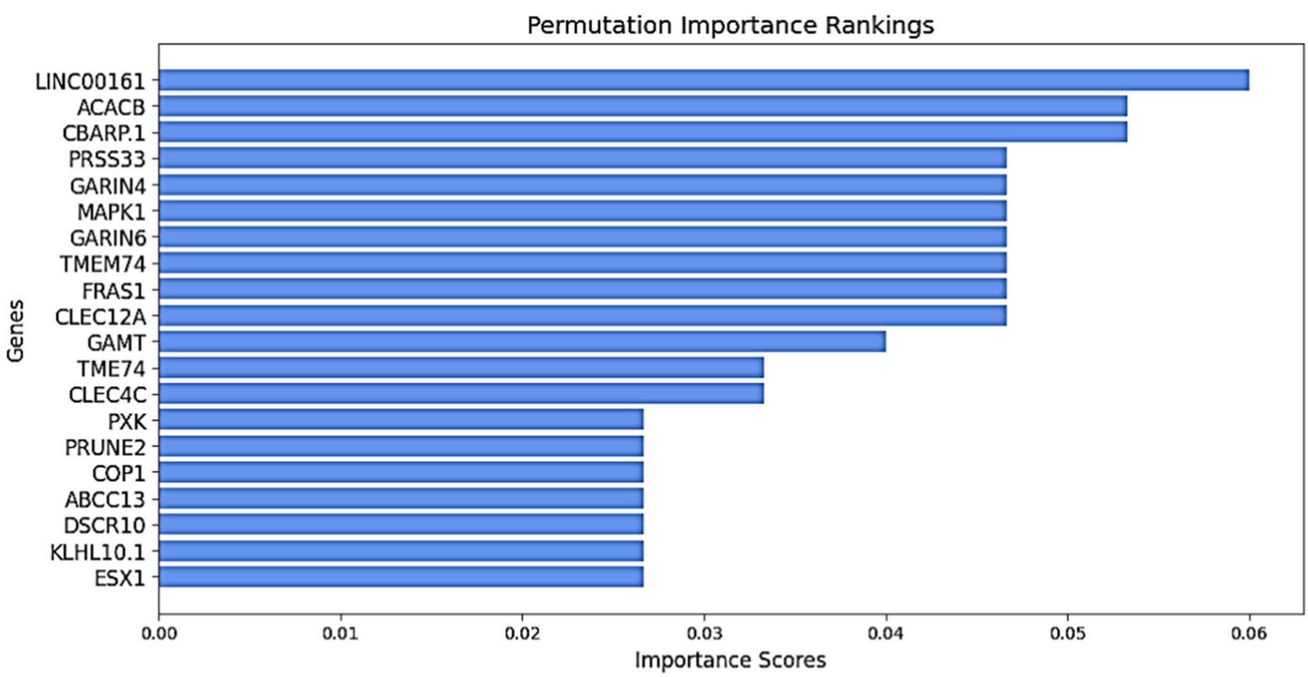

Fig. 3 Bar graph of the top 20 genes of importance

example where these conditions are met, the probability that the predicted class is correct is 100%. However, there are only 0.03% of instances where these conditions are met. That is, the examples where these conditions apply are very limited.

4 Discussion

This research aims to identify potential gene expression markers that can be used to diagnose bladder cancer. For this purpose, the data set downloaded from the GEO database containing gene expression profiles of bladder cancer patients was used. Various XAI and PFI techniques

Table 2 Importance weight values of the first 20 genes of importance

Genes	Weights
LINC00161	0.06
ACACB	0.053
CBARP.1	0.053
PRSS33	0.047
GARIN4	0.047
MAPK1	0.047
GARIN6	0.047
TMEM74	0.047
FRAS1	0.047
CLEC12A	0.047
GAMT	0.04
TME74	0.033
CLEC4C	0.033
PXK	0.027
PRUNE2	0.027
COP1	0.027
ABCC13	0.027
DSCR10	0.027
KLHL10.1	0.027
ESX1	0.027

have been applied to distinguish between tumour and tumour-free classes in this dataset. We use XAI and PFI techniques to understand and explain the decision processes of AI algorithms. These techniques can help us identify potential gene expression markers that can be used to diagnose bladder cancer. In this research, we selected important genes to distinguish between tumour and non-tumour classes using different XAI and PFI techniques. These genes may provide helpful information for the prognosis and treatment of bladder cancer patients. Also, thanks to these techniques, we can understand why algorithms give a particular result and use it safely and accurately in the clinic.

However, XAI also has some ethical, bias and limitation issues. The accuracy and reliability of the explanations given by XAI systems depend on the design of the system and the data used. Therefore, XAI systems must follow ethical principles and explain users' needs appropriately. The ethical problems that may arise in the use of XAI in cancer diagnosis through gene expression profiles include confidentiality of private data, transparency of data collection and processing methods, and unbiasedness of algorithms. Failure to provide sufficient information about

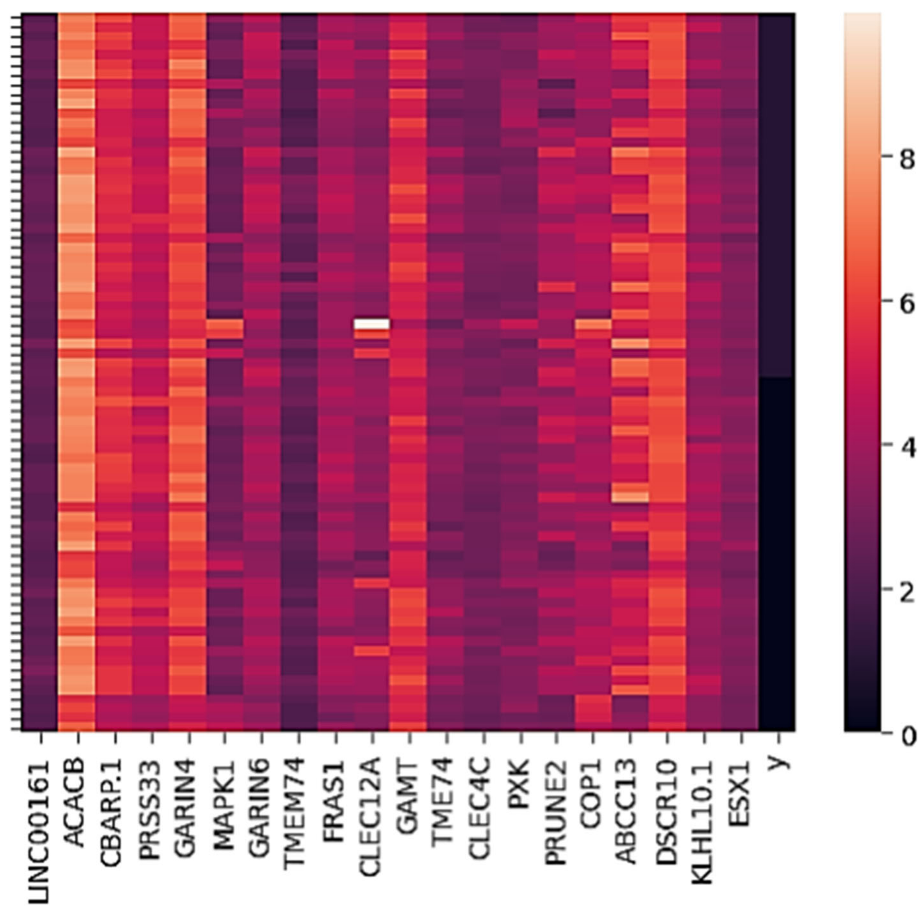

Fig. 4 Heatmap map of the 20 most essential genes obtained using the PFI method

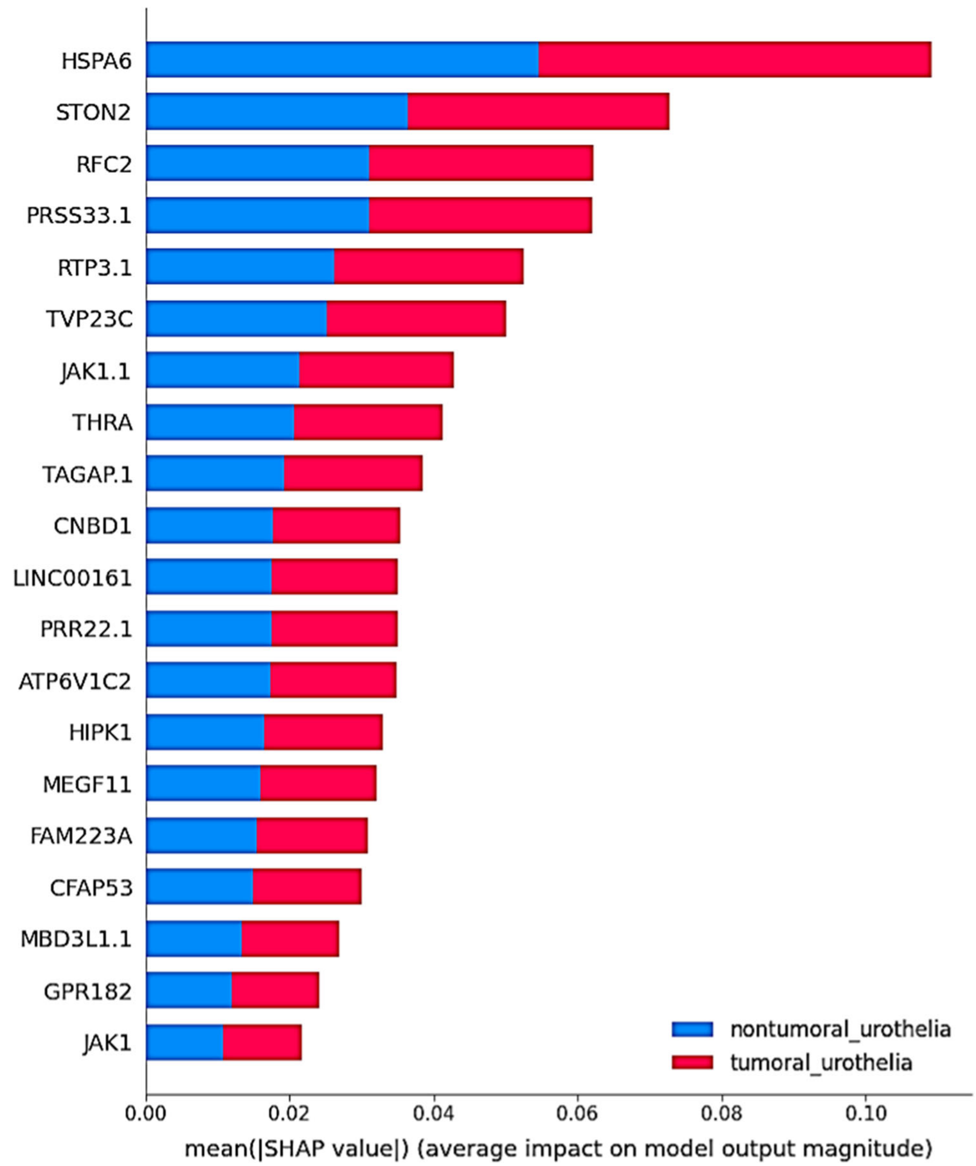

Fig. 5 Summary plot based on Shapley's explanations of 20 genes of importance

the source and processing methods of data used by an XAI system may call into question the reliability and ethics of the system. Therefore, prediction models built on a single data source have many limitations. Among the limitations of using XAI in diagnosing cancer through gene expression profiles, there may be issues such as the system's accuracy and reliability and the explanations' clarity and adequacy. For example, the accuracy and reliability of the explanations given by an XAI system depend on the design of the system and the data used. In addition, the clarity and adequacy of the explanations may vary according to the knowledge level and needs of the users. Therefore, XAI systems must be continuously developed and updated to produce accurate and reliable results. More research and discussion is required to solve these problems. In this research, we would like to state that we are aware of these problems and are trying to solve them.

As a result of the study determined LINC00161, ACACB, and CBARP genes as the genes that best differentiated the tumour class according to the PFI results. Analysis of this dataset provides a perspective for identifying genes that may be effective in bladder cancer classification. The obtained results reflect the importance of genes in classification through weight scores. In particular, genes such as LINC00161, ACACB and CBARP stand out with their high weight scores, indicating that these genes may play a critical role in determining the cancer status of patients. Also, genes with similar weight scores (e.g. PRSS33, GARIN4, MAPK1, GARIN6, TMEM74, FRAS1, etc.) provide further insight into their contribution to disease classification. However, GAMT, TME74, CLEC4C

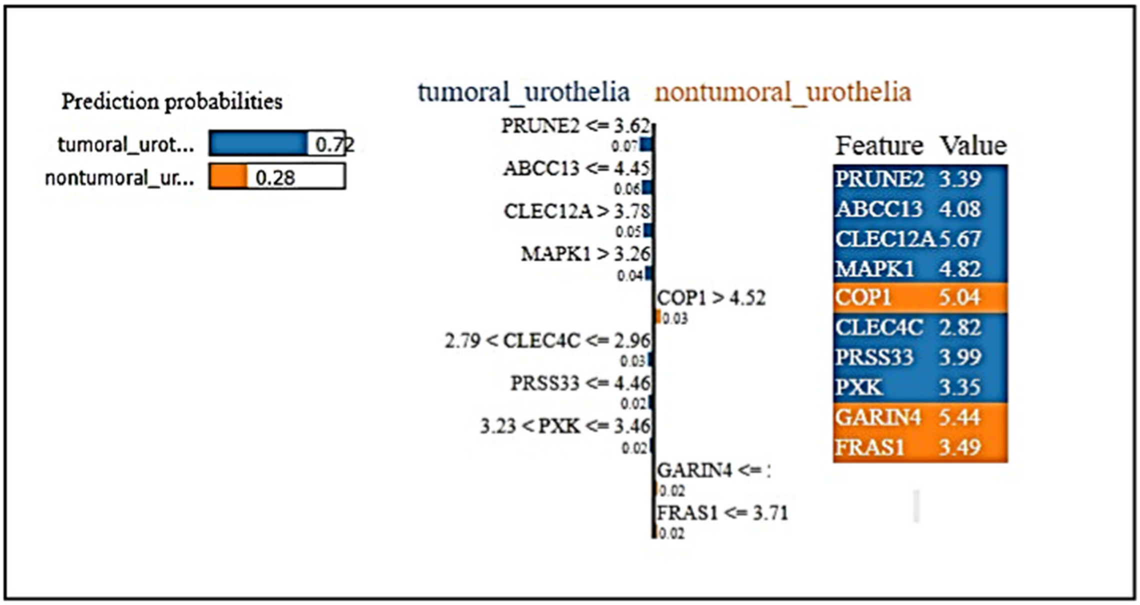

Fig. 6 Visualisation of the effect of different genes involved in bladder cancer on the prediction of bladder cancer by the LIME method

and other genes with lower weight scores should be considered as potential factors. When gene names are examined, it is seen that this gene list represents various biological processes. This may highlight the complexity of the biological mechanisms underlying cancer classification. Therefore, weight scores may not simply reflect the role of genes alone; Interactions between genes and biological context also need to be considered. As a result, the obtained results can guide in determining the focal points for future research. In particular, the genes with the highest and lowest weight scores may deserve further examination to provide further illuminating information on cancer biology and treatment strategies.

HSPA6, STON2, and RFC2 genes selected by the SHap method and PRUNE2 and ABCC13 genes determined by the LIME method were also essential. The Anchor method, on the other hand, gave TMEM74, KLHL10, and GAMT genes as significant output. These results are critical for examining the differences between different XAI methods in identifying potential gene markers that can be used to diagnose bladder cancer. In our study, other genes emerged due to applying different XAI methods. The reasons for these differences can be quite complex and not fully understood. However, several possible explanations and debates exist about the reasons for these differences. First, it should be noted that each method operates under a different assumption. For example, the PFI method tries to identify essential features by calculating a separate feature importance score for each gene in the data set. The SHap method attempts to calculate the contribution of a feature value to a particular prediction result.

On the other hand, the LIME method creates a small model describing the prediction result and estimates feature importance from this model. Also, feature importances can be calculated differently because different ways may have different hyperparameter settings. For example, depending on the model chosen for the LIME method, the importance of certain features may be further emphasised or reduced. As a result, different XAI and PFI methods can often yield different results, and the reasons for these results can be quite complex. Further research may be needed to determine the reasons for these differences. The fact that the UCA1 gene does not appear in the first 20 genes is one of the most important indicators of this situation. It is known that long non-coding RNAs (lncRNA) were initially believed to be only transcriptional noise. Still, some lncRNAs regulate gene expression during chromatin modification, transcription, and post-transcriptional processing steps [55, 56].

Many cancer-associated lncRNAs have been reported to be new independent biomarkers for cancer diagnosis and prognosis in different cancer types, such as breast cancer, oesophagal squamous cell carcinoma, colorectal cancer, lung cancer, and ovarian cancer [57–62]. Some studies have revealed the effects linc00161 on hepatocellular carcinoma cell metastasis [63]. Some studies have shown that linc00161 is upregulated during cisplatin treatment in osteosarcoma cells; overexpression of linc00161 accelerates cisplatin-induced apoptosis and reduces chemotherapy resistance [64, 65]. In a study examining the molecular mechanism of how garlic extract (GE) inhibits the response of bladder cancer EJ cells, Heat shock protein A6 (HSPA6) was chosen as the most upregulated gene responsible for its inhibitory effects, based on microarray datasets. Overexpression of the HSPA6 gene has been found to enhance the GE inhibitory effect of proliferation, migration, and invasion of EJ cells. This enhancing effect of HSPA6 is enhanced by phosphorylation of the ATM-CHK2-Cdc25C-p21WAF1-Cdc2 cascade, MAPK and AKT signalling, and suppression of transcription factor-associated MMP-9 regulation in EJ cells via the G2/M phase in response to GE. These results have shown that HSPA6 may provide a new approach to treating malignancies [66]. Prostate cancer antigen 3 (PCA3) is a unique regulatory mechanism that regulates PRUNE2 levels. This mechanism involves forming a PRUNE2/PCA3 double-stranded RNA, which undergoes adenosine deaminase acting on RNA (ADAR)-dependent adenosine-to-inosine RNA editing. The results of this study show that PCA3 is a dominant-negative oncogene and PRUNE2 is an unrecognised tumour suppressor gene in human prostate cancer [67]. Prostate and bladder cancer are two different types of cancer and are not related to each other. However, they can sometimes be confused with each other as they show similar symptoms. Prostate cancer can cause difficulty urinating, frequent urination, blood in the urine, and changes in sexual function. TMEM74 is a protein that ensures the survival of tumour cells by inducing autophagy. High levels of expression of TMEM74 significantly shorten the survival of patients in many types of cancer [68, 69]. The results of our study seem to fit our research question and hypothesis. The classification model we develop can distinguish between types of bladder cancer with high accuracy, and model agnostic methods used to interpret the model's decisions can reveal bladder cancer-associated genes, their biological significance, and their prognostic value. This makes an important and valuable contribution to bladder cancer diagnosis, prognosis, and improvement of treatment options. Thanks to this study, supported by literature studies, it should be considered that different genes should be studied in bladder cancer and that the surprise gene may be associated with it.

5 Conclusion

The study's main purpose is to use tumoral and non-tumour gene expression data to classify bladder cancer patients. In this context, analyses were made on Gene Expression Omnibus (GEO) and Curated Microarray Database (CuMiDa) data with Permutation Feature Importance (PFI), SHapley Additive ExPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) and Anchor methods. Findings, LINC00161, ACACB and CBARP by PFI method; HSPA6, STON2 and RFC2 according to the SHAP method; The LIME method showed that genes such as PRUNE2 and ABCC13 and TMEM74 were essential. In addition, KLHL10 and GAMT genes were determined by the Anchor method. This study emphasizes that genes detected in other cancer types may also be effective in bladder cancer. In addition, the use of explainable methods in cancer data has been shown to have significant potential to support disease prognosis and treatment strategies in the clinical setting. The main results of the study include LINC00161, ACACB and CBARP determined by the PFI method; HSPA6, STON2 and RFC2 detected by the SHAP method; it is seen that genes such as PRUNE2 and ABCC13 and TMEM74 detected by the LIME method come to the fore. In addition, the study also shows that genes identified in other cancer types may also play an essential role in the field of bladder cancer. It is also among the results that the use of explainable methods in cancer data can evaluate the prognosis of the disease and support the treatment processes of clinical practice. Among the study's strengths is the view that explainable methods can support prognosis and treatment strategies in clinical practice. However, the study's limitations include using only two data sources. Future research may aim to proceed with a broader perspective by considering these recommendations. In this context, areas such as integrating different genomic data sources, discovering new explainability methods, biological validation and combination therapies can be examined. Data mining and ML techniques can also be considered valuable tools for more effective analysis of gene expression data.

Data availability Data openly available in a public repository (https://www.ncbi.nlm.nih.gov/geo/), (https://sbcb.inf.ufrgs.br/cumida).

References

1. WHO. Bladder cancer. https://www.iarc.who.int/cancer-type/bladder-cancer/ (accessed 2023).
2. Segundo-Val IS, Sanz-Lozano CS (2016) Introduction to the gene expression analysis. Methods Mol Biol 1434:29–43. https://doi.org/10.1007/978-1-4939-3652-6_3
3. Vadapalli S, Abdelhalim H, Zeeshan S, Ahmed Z (2022) Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. Brief Bioinf 23(5):191. https://doi.org/10.1093/bib/bbac191
4. Abbas M, El-Manzalawy Y (2020) Machine learning based refined differential gene expression analysis of pediatric sepsis. BMC Med ical Genom 13(1):122. https://doi.org/10.1186/s12920-020-00771-4
5. Guneri-Sozeri PY, Erkek-Ozhan S (2022) Identification of the gene expression changes and gene regulatory aspects in ELF3 mutant bladder cancer. Mol Biol Rep 49(4):3135–3147. https://doi.org/10.1007/s11033-022-07145-2
6. Zaravinos A, Lambrou GI, Volanis D, Delakas D, Spandidos DA (2011) Spotlight on differentially expressed genes in urinary bladder cancer. PLoS ONE 6(4):e18255. https://doi.org/10.1371/journal.pone.0018255

7. Khalsan M et al (2022) A survey of machine learning approaches applied to gene expression analysis for cancer prediction. IEEE Access 10:27522–27534. https://doi.org/10.1109/ACCESS.2022.3146312
8. Rukhsar L, Bangyal WH, Ali Khan MS, Ag Ibrahim AA, Nisar K, Rawat DB (2022) Analyzing RNA-Seq gene expression data using deep learning approaches for cancer classification. Appl Sci 12(4):1850. https://doi.org/10.3390/app12041850
9. Almarzouki HZ (2022) Deep-learning-based cancer profiles classification using gene expression data profile. J Healthcare Eng 2022:4715998. https://doi.org/10.1155/2022/4715998
10. Chen K et al (2021) Identification and validation of hub genes associated with bladder cancer by integrated bioinformatics and experimental assays. Front Oncol Original Res 11:782981. https://doi.org/10.3389/fonc.2021.782981
11. Wagner A (2022) AI predicts the effectiveness and evolution of gene promoter sequences. Nature 603:384. https://doi.org/10.1038/d41586-022-00384-0
12. Abbod MFMF et al. (2006) Artificial intelligence technique for gene expression profiling of urinary bladder cancer. In: 2006 3rd International IEEE conference intelligent systems, 4–6 Sept 2006, pp 646–651. https://doi.org/10.1109/IS.2006.348495
13. Altmann A, Toloşi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. Bioinformatics 26(10):1340–1347
14. Li J et al (2023) Identification of genes related to immune enhancement caused by heterologous ChAdOx1-BNT162b2 vaccines in lymphocytes at single-cell resolution with machine learning methods. Front Immunol 14:1131051. https://doi.org/10.3389/fimmu.2023.1131051
15. Shew M et al (2021) MicroRNA profiling as a methodology to diagnose Ménière's disease: potential application of machine learning. Otolaryngol Head Neck Surg 164(2):399–406. https://doi.org/10.1177/0194599820940649
16. Bazaga A, Leggate D, Weisser H (2020) Genome-wide investigation of gene-cancer associations for the prediction of novel therapeutic targets in oncology. Sci Rep 10(1):10787. https://doi.org/10.1038/s41598-020-67846-1
17. Shapley L (1953) A value for n-person games. Princeton University Press, Princeton, pp 307–317. https://doi.org/10.1515/9781400881970-018
18. Derks J, Peters H (1993) A shapley value for games with restricted coalitions. Int J Game Theory 21(4):351–60. Available: https://EconPapers.repec.org/RePEc:spr:jogath:v:21:y:1993:i:4:p:351-60.
19. Sanchez K, Kamal K, Manjaly P, Ly S, Mostaghimi A (2023) Clinical application of artificial intelligence for non-melanoma skin cancer. Current Treatment Options Oncol 24(4):373–379. https://doi.org/10.1007/s11864-023-01065-4
20. Kumar S, Das A (2023) Peripheral blood mononuclear cell derived biomarker detection using eXplainable Artificial Intelligence (XAI) provides better diagnosis of breast cancer. Comput Biol Chem 104:107867. https://doi.org/10.1016/j.compbiolchem.2023.107867
21. Zhu K et al (2022) A novel 10-gene ferroptosis-related prognostic signature in acute myeloid leukemia. Front Oncol 12:1023040. https://doi.org/10.3389/fonc.2022.1023040
22. Palatnik de Sousa I, Maria Bernardes Rebuzzi Vellasco M, Costa da Silva E (2019) Local interpretable model-agnostic explanations for classification of lymph node metastases. Sensors 19(13), 2969. Available: https://www.mdpi.com/1424-8220/19/13/2969
23. Lai Y et al (2022) Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning. Front Immunol 13:1046410. https://doi.org/10.3389/fimmu.2022.1046410
24. Oni O, Qiao S (2019) Model-agnostic interpretation of cancer classification with multi-platform genomic data, pp 34–41
25. Modhukur V et al. (2021) Machine learning approaches to classify primary and metastatic cancers using tissue of origin-based DNA methylation profiles. Cancers (Basel) 13(15):3768. Available: https://www.mdpi.com/2072-6694/13/15/3768
26. Marco Tulio Ribeiro SS, Guestrin C (2018) Anchors: high-precision model-agnostic explanations. . Available: https://homes.cs.washington.edu/~marcotcr/aaai18.pdf
27. Edgar R, Domrachev M, Lash AE (2002) "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30(1):207–210. https://doi.org/10.1093/nar/30.1.207
28. Feltes BC, Chandelier E, Grisci B, Dorn M (2019) CuMiDa: an extensively curated microarray database for benchmarking and testing of machine learning approaches in cancer research. J Comput Biol 26. https://doi.org/10.1089/cmb.2018.0238
29. Sherman BT et al (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res 50(W1):W216-w221. https://doi.org/10.1093/nar/gkac194
30. Botchkarev A (2018) Performance metrics (error measures) in machine learning regression. Forecast Prognost Prop Typol
31. Vujovic ZD (2021) Classification model evaluation metrics. Int J Adv Comput Sci Appl 12(6):599–606
32. De Diego IM, Redondo AR, Fernandez RR, Navarro J, Moguerza JM (2022) General performance score for classification problems. Appl Intell 52(10):12049–12063. https://doi.org/10.1007/s10489-021-03041-7
33. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/a:1010933404324
34. Octaviani TL, Rustam Z (2019) Random forest for breast cancer prediction. In: 4th International symposium on current progress in mathematics and sciences (ISCPMS). Univ Indonesia, Fac Math and Nat Sci, Depok, INDONESIA, vol 2168. In: AIP Conference Proceedings,30–31 Oct 2018. https://doi.org/10.1063/1.5132477. Available: <Go to ISI>://WOS:000519032600050
35. Huljanah M, Rustam Z, Utama S, Siswantining T, Iop (2019) Feature selection using random forest classifier for predicting prostate cancer. In: presented at the 9TH annual basic science international conference 2019 (BASIC 2019)
36. Huang M et al (2017) Head and neck cancer survival outcome prediction based on NRG oncology RTOG 0522 with random forests and random survival forests. Med Phys 44(6)
37. Liu DF et al (2021) Optimisation and evaluation of the random forest model in the efficacy prediction of chemoradiotherapy for advanced cervical cancer based on radiomics signature from high-resolution T2 weighted images. Arch Gynecol Obstetrics 303(3):811–820. https://doi.org/10.1007/s00404-020-05908-5.
38. Santhanam R, Uzir N, Raman S, Banerjee S (2017) Experimenting XGBoost algorithm for prediction and classification of different datasets
39. Deng XS, Li M, Deng SB, Wang L (2022) Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification. Med Biol Eng Comput 60(3):663–681. https://doi.org/10.1007/s11517-021-02476-x
40. Ma BS et al (2022) Diagnostic classification of cancers using DNA methylation of paracancerous tissues. Sci Rep 12(1):10646. https://doi.org/10.1038/s41598-022-14786-7
41. Song YY, Lu Y (2015) Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 27(2):130–135. https://doi.org/10.11919/j.issn.1002-0829.215044
42. Zhang Z (2016) Introduction to machine learning: k-nearest neighbors. Ann Transl Med 4(11):218. https://doi.org/10.21037/atm.2016.03.37

43. Lee WM (2019) Supervised learning—classification using K-nearest neighbors (KNN), pp 205–220
44. Momodu A (2017) K-nearest neighbor implementation in python 3.6.1 from scratch
45. Gao S, Li HM (2012) IEEE breast cancer diagnosis based on support vector machine. In: Presented at the 2012 2nd international conference on uncertainty reasoning and knowledge engineering (URKE)
46. Chen LY, Li JT, Chang MM (2020) Cancer diagnosis and disease gene identification via statistical machine learning. Curr Bioinform 15(9):956–962. https://doi.org/10.2174/1574893615666200207094947
47. Teeyapan K, Theera-Umpon N, Auephanwiriyakul S, IEEE (2015) Application of support vector based methods for cervical cancer cell classification. In: Presented at the proceedings 5th IEEE international conference on control system, computing and engineering (ICCSCE 2015)
48. Liu TB, Zhang XM, Chen R, Deng XX, Fu B (2023) Development, comparison, and validation of four intelligent, practical machine learning models for patients with prostate-specific antigen in the gray zone. Front Oncol 13. Art no. 1157384. https://doi.org/10.3389/fonc.2023.1157384
49. Akcay M, Etiz D, Celik O, Ozen A (2022) Evaluation of acute hematological toxicity by machine learning in gynecologic cancers using postoperative radiotherapy. Indian J Cancer 59(2):178–186. https://doi.org/10.4103/ijc.IJC_666_19
50. Lei L, IEEE (2018) Research on logistic regression algorithm of breast cancer diagnose data by machine learning. In: presented at the 2018 international conference on robots and intelligent system (ICRIS 2018)
51. Ramirez SG, Hales RC, Williams GP, Jones NL (2022) Extending SC-PDSI-PM with neural network regression using GLDAS data and Permutation Feature Importance. Environ Model Softw 157:105475
52. Gramegna A, Giudici P (2021) SHAP and LIME: an evaluation of discriminative power in credit risk. Front Artif Intell 4. Art no. 752558. https://doi.org/10.3389/frai.2021.752558
53. Holzinger A, Saranti A, Molnar C, Biecek P, Samek W (2022) Explainable AI methods—a brief overview. Springer International Publishing, pp 13–38
54. Hagras H (2018) Toward human-understandable, explainable AI. Computer 51(9):28–36
55. Shi Y, Zhou Y (2010) The role of surgery in the treatment of gastric cancer. J Surg Oncol 101(8):687–692. https://doi.org/10.1002/jso.21455
56. Wilusz JE, Sunwoo H, Spector DL (2009) Long non-coding RNAs: functional surprises from the RNA world. Genes Dev 23(13):1494–1504. https://doi.org/10.1101/gad.1800909
57. Shen Y et al (2015) Prognostic and predictive values of long non-coding RNA LINC00472 in breast cancer. Oncotarget 6(11):8579–8592. https://doi.org/10.18632/oncotarget.3287
58. Sun J et al (2015) A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. Sci Rep 5(1):16553. https://doi.org/10.1038/srep16553
59. Li J et al (2014) LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut 63(11):1700–1710. https://doi.org/10.1136/gutjnl-2013-305806
60. Hu Y et al (2014) A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. Oncotarget 5(8):2230–2242. https://doi.org/10.18632/oncotarget.1895
61. Zhou M et al (2015) A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. J Trans Med 13(1):231. https://doi.org/10.1186/s12967-015-0556-3
62. Zhou M et al (2016) Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. Oncotarget 7(22):32433–32448. https://doi.org/10.18632/oncotarget.8653
63. Xu LC et al (2017) Up-regulation of LINC00161 correlates with tumor migration and invasion and poor prognosis of patients with hepatocellular carcinoma. Oncotarget 8(34):56168–56173. https://doi.org/10.18632/oncotarget.17040
64. Li Z, Dou P, Liu T, He S (2017) Application of long non-coding RNAs in osteosarcoma: biomarkers and therapeutic targets. Cell Physiol Biochem 42(4):1407–1419. https://doi.org/10.1159/000479205
65. Wang Y et al (2016) Long non-coding RNA LINC00161 sensitises osteosarcoma cells to cisplatin-induced apoptosis by regulating the miR-645-IFIT2 axis. Cancer Lett 382(2):137–146. https://doi.org/10.1016/j.canlet.2016.08.024
66. Shin SS et al (2017) HSPA6 augments garlic extract-induced inhibition of proliferation, migration, and invasion of bladder cancer EJ cells; Implication for cell cycle dysregulation, signaling pathway alteration, and transcription factor-associated MMP-9 regulation. PLoS ONE 12(2):e0171860. https://doi.org/10.1371/journal.pone.0171860
67. Salameh A et al (2015) PRUNE2 is a human prostate cancer suppressor regulated by the intronic long non-coding RNA PCA3. Proc Natl Acad Sci USA 112(27):8403–8408. https://doi.org/10.1073/pnas.1507882112
68. Zhou C, Li AH, Liu S, Sun H () Identification of an 11-autophagy-related-gene signature as promising prognostic biomarker for bladder cancer patients. Biology (Basel) 10(5). https://doi.org/10.3390/biology10050375
69. Sun Y et al (2017) TMEM74 promotes tumor cell survival by inducing autophagy via interactions with ATG16L1 and ATG9A. Cell Death Dis 8(8):e3031. https://doi.org/10.1038/cddis.2017.370