Early diagnosis of pancreatic cancer by machine learning methods using urine biomarker combinations

İREM ACER
FIRAT ORHAN BULUCU
SEMRA İÇER
FATMA LATİFOĞLU

Follow this and additional works at: https://journals.tubitak.gov.tr/elektrik

Part of the Computer Engineering Commons, Computer Sciences Commons, and the Electrical and Computer Engineering Commons

Recommended Citation
ACER, İREM; BULUCU, FIRAT ORHAN; İÇER, SEMRA; and LATİFOĞLU, FATMA (2023) "Early diagnosis of pancreatic cancer by machine learning methods using urine biomarker combinations," Turkish Journal of Electrical Engineering and Computer Sciences: Vol. 31: No. 1, Article 8. https://doi.org/10.55730/1300-0632.3974
Available at: https://journals.tubitak.gov.tr/elektrik/vol31/iss1/8

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Electrical Engineering and Computer Sciences by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.
Early diagnosis of pancreatic cancer by machine learning methods using urine biomarker combinations

İrem ACER1•, Fırat ORHANBULUCU2•,3•, Semra İÇER3•,6•, Fatma LATİFOĞLU3•
1Department of Biomedical Device Technology, Kütahya Dumlupınar University, Kütahya, Turkey
2Department of Biomedical Engineering, İnönü University, Malatya, Turkey
3Department of Biomedical Engineering, Erciyes University, Kayseri, Turkey

Abstract: The most common type of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC), which accounts for the vast majority of pancreatic cancers. The five-year survival rate for PDAC due to late diagnosis is 9%. Early diagnosed PDAC patients survive longer than patients diagnosed at a more advanced stage. Biomarkers can play an essential role in the early detection of PDAC to assist the health professional. Machine learning and deep learning methods are used with biomarkers obtained in recent studies for diagnostic purposes. In order to increase the survival rates of PDAC patients, early diagnosis of the disease with a noninvasive test is a critical need. Our study offers a promising approach for the early detection of PDAC with noninvasive urinary biomarkers and carbohydrate antigen 19-9 (CA19-9). The Kaggle Urinary Biomarkers for Pancreatic Cancer (2020) open-access dataset consisting of 590 participants was used in this study. Seven machine learning classifiers (support vector machine (SVM), naive Bayes (NB), k-nearest neighbors (kNN), random forest (RF), light gradient boosting machine (LightGBM), AdaBoost, and gradient boosting classifier (GBC)) to detect PDAC disease classifier were used. Binary and multiple classification processes were carried out. Data was validated in our study using 5–10-fold crossvalidation. This study aimed to determine the best machine learning model by analyzing the performance of machine learning models in determining the classes of healthy controls, pancreatic disorders, and patients with PDAC. It is a remarkable finding that ensemble learning models were more successful in all our groups. The most successful classification method in classifying healthy controls and patients with PDAC was CV-10, while the GBC (92.99%) model was (AUC = 0.9761). The most successful classification method in classifying patients with pancreatic disorders and PDAC was CV-10, while the LightGBM (86.37%) model was (AUC = 0.9348). In the classification of healthy controls, pancreatic disorders, and patients with PDAC, the most successful classification method was CV-5, while the GBC (72.91%) model was (AUC = 0.8733).

Key words: Pancreatic cancer, urine biomarker, machine learning, ensemble learning, classification

1. Introduction
The importance of cancer is increasing day by day due to aging and population growth worldwide [1]. The most common type of pancreatic cancer is PDAC, which accounts for the vast majority of pancreatic cancers [2]. The mortality rate is high because it is usually diagnosed when the disease is advanced. The five-year survival rate for PDAC due to late diagnosis is 9% [3]. PDAC is currently the third leading cause of cancer-related death in the United States [4]. In the study by Rahib et al., it is estimated that pancreatic cancer will be the second most lethal type of cancer worldwide by 2030 [5].
Recent studies have focused on identifying specific biomarkers for PDAC [6–8]. Biomarkers can play an essential role in the early detection of PDAC to assist the health professional. CA19-9 is currently the most routinely used serum biomarker for PDAC. Nevertheless, it does not have the necessary accuracy to detect PDAC by itself [9]. Biomarkers that could potentially improve the earlier detection of PDAC in combination or alone with CA19-9 of the biomarkers examined in studies were mentioned [6, 10]. Advances in data analysis give good results in classifying cancer types using artificial intelligence. Artificial intelligence applications in pancreatic cancer can improve early diagnosis, treatment, and survival rates [11]. Diagnosis of pancreatic cancer has mostly been made through personal health records and imaging methods [12–14]. Machine learning and deep learning methods are used with biomarkers obtained in the recent studies for diagnostic purposes. Almeida et al. created an artificial neural network model to predict PDAC based on gene expression. They classified network tumor samples with an f1 score of 0.83 for normal samples and 0.88 for PDAC samples [11]. Honda et al. correctly diagnosed 97.2% of cancer patients in the training cohort and 94.4% of healthy controls using the SVM model based on four plasma proteins. In the test set, it was detected 91% correctly [15]. Hsieh et al. developed a prediction model for pancreatic cancer in patients with type 2 diabetes using logistic regression and artificial neural network models. As a result of their studies, they stated that the logistic regression (LR), (AUC = 0.727) model was more successful than the artificial neural network (AUC = 0.605) model [16]. In the study of Khatri et al., gene-based PDAC classifier performance using a support vector machine showed a sensitivity (Sens.) of 0.88 and a specificity (Spec.) of 0.80 for blood datasets, a sensitivity of about 0.94 and a specificity of 0.97 for tissue datasets [17]. Radon et al. analyzed urine samples and identified three biomarkers for early detection of radon PDAC [8]. They then examined a new biomarker with additional urine samples collected [18]. They stated that substitution of REG1B (Sens. and Spec. >85%) biomarker instead of REG1A biomarker (LYVE-1+REG1A+TFF1 Sens. = 76.9%, Spec. = 89.8%) increased the detection performance of resectable PDAC. Their study applied logistic regression to CA19-9 and its combinations (LYVE1, REG1, TFF1, creatinine, and age) by separating training and test data in a 1:1 ratio. The success of urinary biomarkers in the study was confirmed with many samples.

In order to increase the survival rates of PDAC patients, early diagnosis of the disease with a noninvasive test is a critical need. Our study offers a promising approach for the early detection of PDAC with noninvasive urinary biomarkers and CA19-9. To the best of our knowledge, no comparative classification studies have been conducted using various machine learning models with urine biomarkers for PDAC detection. The use of machine learning methods in PDAC patients can increase early diagnosis and survival. This study aimed to accurately diagnose PDAC patients using urine biomarkers and various machine learning methods. The main outputs of this study can be listed as follows:

1. To detect PDAC disease, binary, and multiple classifications were made with seven machine learning models.
2. The classification model was designed using SVM, NB, KNN, RF, LightGBM, AdaBoost, and GBC for healthy controls, pancreatic disorders, and PDAC patients.
3. While the most successful result in the classification of the three groups is CV-5, it is seen that the GBC (72.91%) classifier is more successful than the other models according to the accuracy rate metric.
4. While the most successful result is CV-10 in classifying healthy controls and patients with PDAC, the GBC (92.99%) classifier seems to be more successful than other models according to the accuracy rate metric.
5. While the most successful result is CV-10 in classifying patients with pancreatic disorders and patients
with PDAC, it is seen that the LightGBM (86.37%) classifier is more successful than other models according to the accuracy rate metric.

In the second part of the study, information about the data set is given, and the methods mentioned above are explained. In the third chapter, the findings obtained from these methods are presented. In the last chapter, the results obtained in the study are discussed together with the literature.

2. Materials and methods

2.1. Data set

The Kaggle Urinary Biomarkers for Pancreatic Cancer (2020) open-access dataset was used in this study [18]. The data set used consists of 590 participants. The dataset includes three groups: healthy controls, patients with noncancerous pancreatic disorders such as chronic pancreatitis, and patients with pancreatic ductal adenocarcinoma. The distribution of the data used is shown in Table 1.

<table>
<thead>
<tr>
<th>Label</th>
<th>Classes</th>
<th>Number of data</th>
<th>Sex</th>
<th>Mean age (±SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Healthy controls</td>
<td>183</td>
<td>115 Female (F)-68 Male (M)</td>
<td>56.33 (± SD 12.20)</td>
</tr>
<tr>
<td>2</td>
<td>Pancreatic patients</td>
<td>208</td>
<td>101 F-107 M</td>
<td>54.70 (± SD 13.34)</td>
</tr>
<tr>
<td>3</td>
<td>Pancreatic ductal adenocarcinoma</td>
<td>199</td>
<td>83 F-116 M</td>
<td>66.18 (± SD 10.51)</td>
</tr>
</tbody>
</table>

*SD: Standard deviation

Eight features were used in the dataset, including age, sex, five urinary biomarkers, and the FDA-approved CA19-9 plasma biomarker. The graph showing the correlation between the features is shown in Figure 1. Figure 1 expresses the degree of dependence between a variable used in the study and every other variable. To best generalize the dataset, it is desired that the dataset covers most of the feature space. This means the selection of dissimilar features. As seen in Figure 1, the features used in the study were chosen in a way to generalize the data set, that is, features that were not similar to each other were selected. According to the color scale, the correlation relationship between the columns close to blue is high, while the correlation relationship decreases gradually in the colors toward red.

The key features are four urinary biomarkers: LYVE1, REG1B, REG1A, TFF1 and urine creatinine.
- Creatinine is a protein often used as an indicator of kidney function.
- LYVE1 is lymphatic vessel endothelial hyaluronan receptor 1, a protein that may play a role in tumor metastasis.
- The REG1A (regenerating islet-derived 1 alpha) gene product, a member of the REG glycoprotein family, has been associated with pancreatitis [8, 19].
- REG1B (regenerating islet-derived 1 beta) is a protein that may be associated with pancreatic regeneration [20].
- TFF1 is trefoil factor 1, which plays a role in protecting epithelial cells and the regeneration and repair of the urinary tract [21].

2.2. Classifications

Classification is a supervised learning approach where the program learns from the given data entry and then uses this learning to classify new observations according to their characteristics. The increase in medical data
also increases the need for effective classifiers in the field of machine learning. In this study, healthy controls-PDAC patients, patients with pancreatic disease-PDAC patients, and all three groups were classified using the most widely used machine learning algorithms (SVM, KNN, NB) and ensemble learning classifiers (RF, GBC, Light GBM, AdaBoost). Classification processes were carried out using the Python programming language. Traditional algorithms such as KNN, SVM, and NB have a lower performance, while ensemble algorithms have a relatively higher accuracy. As a preprocessing step, missing values in the data were filled with the mean. Data was validated in our study using 5–10-fold crossvalidation. The methodology applied in our study is explained in Figure 2. Information about the classifiers used in the study is as follows:

**SVM:** It was first proposed by Cortes and Vapnik in 1995 [22]. It can be used in classification and regression problems. Its development in both practical and theoretical fields is related to its solid mathematical background, widely accessible software applications, and ease of use [23]. Its purpose is to find a hyperplane that can separate the class of given features with a maximal margin between the categories. The SVM algorithm
steps in on how to draw this boundary [24]. The line’s location is intended to be at the furthest point between
the two groups, allowing the two classes to be separated as widely as possible.

KNN: It was first proposed by Fix and Hodges in 1951. It is a simple and effective method for
classification [25]. It finds the nearest neighbors among the variables for classification. The most important
point is the distance between the data points. k is a parameter used to determine distances. To classify the
data set, there are nearest neighbors between the variables [26].

NB: It was introduced by Thomas Bayes in 1812 [27]. The algorithm calculates the probability of each
situation for an element and classifies it according to the highest probability value. Very successful classifications
can be made with a small number of training data.

Ensemble learning can provide more consistent and accurate predictions than a single prediction model.
It is a current field of machine learning research that first trains a set of individual learners and then combines
them with some strategies to improve overall performance. The goal is to improve accuracy and significantly
reduce classification errors by combining estimates [28, 29]. In this study, RF, Light GBM, AdaBoost and GBC
were used as ensemble learning classifiers.

RF: RF is one of the best known, widely used techniques for classification in a variety of applications.
It is an ensemble learning method consisting of several decision tree classifiers to obtain the optimal solution
and increase the classification value [30].

Light GBM: LightGBM is a decision tree-based ensemble learning method which is a vertical tree
growth procedure. It generalizes well and combines the predictions of multiple decision trees to make the final
prediction [31].

AdaBoost: In the AdaBoost algorithm, the training set is first trained with a weak learner. At each
posttraining stage, it is retrained by rerunning the classifier by increasing the weight of the wrong predictions
made as a result of the previous stage. With these processes, it is aimed to focus on the wrong predictions and
to increase the accuracy rate of the created model in classification [32].

GBC: GBC is a decision tree-based estimation algorithm used for classification problems. It uses an
amplification technique that combines a series of weak tutorials to build a strong model. It is specified with a
loss function and a weak base classifier. The goal here is to estimate an aggregate model that minimizes the
loss function [33].

2.3. Performance measures
The performances of the classification algorithms used in the study were calculated according to the confusion
matrix given below in Figure 3 for standard performance criteria such as accuracy (Acc.), recall, precision
(Prec.), and area under the curve (AUC) values.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td>Negative</td>
<td>TN</td>
</tr>
<tr>
<td>Positive</td>
<td>FN</td>
</tr>
</tbody>
</table>

Figure 3. Confusion matrix.
Accuracy means the percentage of samples that make the correct guess among all available samples. The recall is indicated how successfully positive states were predicted. Precision is defined as a corrective prediction in the positive class of samples. In Equations 1 – 3, true positive (TP) means that the true patient class and the class found by the algorithm match correctly, while false positive (FP) means that the nonpatient class is predicted as the patient. Likewise, true negative (TN) means that the class that is not sick is predicted correctly by the algorithm, while a false negative (FN) means that the class that is sick is incorrectly predicted as the class that is not sick. The receiver operating characteristic (ROC) curve is used to describe the accuracy of the diagnostic test itself and to allow a reliable comparison between tests. AUC is an indicator of how well machine learning models can distinguish classes. As the area under the curve increases, the discrimination performance between classes increases. Kappa statistical value is a statistical criterion used to test the reliability of algorithms.

Crossvalidation was set as 5 and 10 in the classification processes and reported according to the highest value. Cohen’s kappa coefficient is a statistical measure used to test the reliability of two raters. It was stated that the closer the kappa statistical value was to 0, the worse the agreement, and the closer to 1, the better the agreement [34].

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100
\]  
\[
\text{Recall} = \frac{TP}{TP + FN} \times 100
\]  
\[
\text{Precision} = \frac{TP}{TP + FP} \times 100
\]

3. Results

In this study, pancreatic cancer detection was performed with SVM, KNN, NB, and ensemble learning (RF, Light GBM, AdaBoost ve GBC) machine learning algorithms on the Kaggle Urinary Biomarkers for Pancreatic Cancer (2020) open-access dataset. The results of the machine learning techniques used are given comparatively. The classification results of the study as healthy controls and patients with pancreatic ductal adenocarcinoma, patients with noncancerous pancreatic disorders such as chronic pancreatitis, and patients with pancreatic ductal adenocarcinoma and 3 groups together are given below. All models were trained with 5- and 10-fold crossvalidation techniques, and results of accuracy, recall, precision, and AUC scores were obtained.

3.1. Classification of healthy controls and patients with PDAC

The evaluation of the results of the classification algorithms used is given in Table 2. It has been seen that ensemble learning algorithms are more successful in classification than classical classification algorithms. In the results, it is seen that the GBC classifier is more successful than other models according to the accuracy rate metric from seven different machine learning models trained with 5- and 10-fold crossvalidation techniques in classifying healthy controls and patients with PDAC. The most successful result was given by the GBC (92.99%) model when CV-10.
3.2. Classification of patients with pancreatic disorders and PDAC

The evaluation of the results of the classification algorithms used is given in Table 3. As in the other group, ensemble learning algorithms were more successful than classical classification algorithms. The results show that the LightGBM classifier is more successful than the other models according to the accuracy rate metric from 7 different machine learning models trained with the 5- and 10-fold crossvalidation technique in the classification of patients with pancreatic cancer disorders and PDAC. The LightGBM (86.37%) model gave the most successful result when CV-10.

<table>
<thead>
<tr>
<th>Model</th>
<th>CV-5 Acc. (%)</th>
<th>CV-5 AUC</th>
<th>CV-5 Recall</th>
<th>CV-5 Prec.</th>
<th>CV-5 Kappa</th>
<th>CV-10 Acc. (%)</th>
<th>CV-10 AUC</th>
<th>CV-10 Recall</th>
<th>CV-10 Prec.</th>
<th>CV-10 Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBC</td>
<td>92.10</td>
<td>0.9734</td>
<td>0.9335</td>
<td>0.9242</td>
<td>0.8408</td>
<td>92.99</td>
<td>0.9761</td>
<td>0.9245</td>
<td>0.9368</td>
<td>0.8598</td>
</tr>
<tr>
<td>LightGBM</td>
<td>92.10</td>
<td>0.9742</td>
<td>0.9213</td>
<td>0.9347</td>
<td>0.8412</td>
<td>92.13</td>
<td>0.9731</td>
<td>0.9134</td>
<td>0.9315</td>
<td>0.8425</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>89.45</td>
<td>0.9620</td>
<td>0.8963</td>
<td>0.9112</td>
<td>0.7879</td>
<td>91.24</td>
<td>0.9615</td>
<td>0.9013</td>
<td>0.9233</td>
<td>0.8248</td>
</tr>
<tr>
<td>RF</td>
<td>89.46</td>
<td>0.9578</td>
<td>0.9029</td>
<td>0.9041</td>
<td>0.7882</td>
<td>90.38</td>
<td>0.9726</td>
<td>0.8961</td>
<td>0.9142</td>
<td>0.8076</td>
</tr>
<tr>
<td>KNN</td>
<td>89.14</td>
<td>0.9292</td>
<td>0.8904</td>
<td>0.9107</td>
<td>0.7816</td>
<td>86.57</td>
<td>0.9252</td>
<td>0.85</td>
<td>0.8812</td>
<td>0.7317</td>
</tr>
<tr>
<td>NB</td>
<td>77.34</td>
<td>0.9218</td>
<td>0.6298</td>
<td>0.9341</td>
<td>0.5569</td>
<td>79.88</td>
<td>0.9274</td>
<td>0.6474</td>
<td>0.9309</td>
<td>0.5987</td>
</tr>
<tr>
<td>SVM</td>
<td>71.12</td>
<td>0.00</td>
<td>0.7092</td>
<td>0.8068</td>
<td>0.4145</td>
<td>68.20</td>
<td>0.00</td>
<td>0.7340</td>
<td>0.7147</td>
<td>0.3626</td>
</tr>
</tbody>
</table>

3.3. Classification of healthy controls, pancreatic disorders, and patients with PDAC

The evaluation of the results of the classification algorithms used is given in Table 4. It was observed that ensemble learning algorithms were more successful in classification than classical classification algorithms in all groups. In the results, it is seen that the GBC classifier is more successful than other models according to the accuracy rate metric from 7 different machine learning models trained with 5 crossvalidation techniques in classifying healthy controls, pancreatic disorders, and patients with PDAC. The GBC model (72.91%) gave the most successful result when CV-5. Figure 4 shows the GBC ROC curve when CV-5, which is the most successful result. In Figure 5, the feature importance graph in this classification is given. According to the accuracy rate metric from 7 different machine learning models trained with 10 crossvalidation techniques, the LightGBC classifier was found to be more successful than other models in the classification of healthy controls, pancreatic disorders, and patients with PDAC.
Table 4. Classification results of healthy controls, pancreatic disorders, and patients with PDAC.

<table>
<thead>
<tr>
<th>Model</th>
<th>CV Acc. (%)</th>
<th>CV-5 AUC</th>
<th>CV-5 Recall</th>
<th>CV-5 Prec.</th>
<th>CV-5 Kappa</th>
<th>CV-10 Acc. (%)</th>
<th>CV-10 AUC</th>
<th>CV-10 Recall</th>
<th>CV-10 Prec.</th>
<th>CV-10 Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBC</td>
<td>72.91</td>
<td>0.8733</td>
<td>0.7788</td>
<td>0.7310</td>
<td>0.5954</td>
<td>70.83</td>
<td>0.8698</td>
<td>0.7108</td>
<td>0.7125</td>
<td>0.5614</td>
</tr>
<tr>
<td>LightGBM</td>
<td>69.31</td>
<td>0.8793</td>
<td>0.6902</td>
<td>0.6983</td>
<td>0.5377</td>
<td>72.33</td>
<td>0.8717</td>
<td>0.7251</td>
<td>0.7302</td>
<td>0.5840</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>61.69</td>
<td>0.7353</td>
<td>0.6170</td>
<td>0.6259</td>
<td>0.4233</td>
<td>55.95</td>
<td>0.7210</td>
<td>0.5594</td>
<td>0.5772</td>
<td>0.3353</td>
</tr>
<tr>
<td>RF</td>
<td>68.68</td>
<td>0.8584</td>
<td>0.6857</td>
<td>0.3890</td>
<td>0.5282</td>
<td>67.82</td>
<td>0.8582</td>
<td>0.6794</td>
<td>0.6831</td>
<td>0.5156</td>
</tr>
<tr>
<td>KNN</td>
<td>64.45</td>
<td>0.7955</td>
<td>0.6471</td>
<td>0.6544</td>
<td>0.4664</td>
<td>62.94</td>
<td>0.7805</td>
<td>0.6330</td>
<td>0.6402</td>
<td>0.4444</td>
</tr>
<tr>
<td>NB</td>
<td>52.35</td>
<td>0.7459</td>
<td>0.5424</td>
<td>0.5752</td>
<td>0.2979</td>
<td>51.79</td>
<td>0.7305</td>
<td>0.5348</td>
<td>0.5733</td>
<td>0.2872</td>
</tr>
<tr>
<td>SVM</td>
<td>43.87</td>
<td>0.00</td>
<td>0.44</td>
<td>0.4233</td>
<td>0.1559</td>
<td>45.79</td>
<td>0.00</td>
<td>0.4677</td>
<td>0.3895</td>
<td>0.1933</td>
</tr>
</tbody>
</table>

Figure 4. Three-class gradient boosting classifier ROC curve (CV-5).

4. Conclusion and discussion

PDAC is the most common type of pancreatic cancer. Diagnosing PDAC at an early stage is very important for increasing the survival rate. It is important to distinguish PDAC from other nonmalignant benign gastrointestinal diseases. Early diagnosed PDAC patients survive longer than patients diagnosed at a more advanced stage. Therefore, it is essential to diagnose the disease at an early stage. The lack of clinically approved biomarkers with the required performance criteria is associated with the lack of early detection of PDAC [35]. Evaluation of biomarkers in body fluids such as blood, tissue, urine, saliva, or pancreatic juice is partially invasive for diagnosing PDAC. However, it can enable early diagnosis and intervention of PDAC and is somewhat less expensive compared to existing methods. Another challenge with early diagnosis is differentiating PDAC from other nonmalignant benign gastrointestinal diseases such as chronic pancreatitis. It is difficult to make a differential diagnosis because the imaging and clinical patterns are generally similar [36, 37]. This study aimed to determine the best machine learning classifier by analyzing the performance of machine learning models in
determining the classes of healthy controls, pancreatic disorders, and patients with PDAC and to contribute to early diagnosis.

Ensemble learning algorithms showed higher success in all classifications than classical classification algorithms. This result parallels with the research result of Gupta et al. [38] on cancer diagnosis. Urine biomarkers used in combination with CA19-9 are a promising approach for earlier detection of PDAC. In one study, it was stated that a combination of urinary biomarkers (LYVE-1, REG1A, and TFF1) could be used in the diagnosis of pancreatic cancer [8]. In a later study, it was stated that using REG1B instead of REG1A increased the detection performance of PDAC [18]. When we look at the feature importance graph in Figure 5, it is seen that REG1B is a more important feature than REG1A for classification.

The most successful classification method in classifying healthy controls and patients with PDAC was CV-10, while the GBC (Acc. = 92.99%) model was (AUC = 0.976). In Figure 6, the results of all classification algorithms according to the accuracy metric are given in the graph.

![Feature Importance Plot](image)

**Figure 5.** Feature importance graph.

The most successful classification method in classifying patients with pancreatic disorders and PDAC was CV-10, while the LightGBM (Acc. = 86.37%) model was (AUC = 0.935). In Figure 7, the results of all classification algorithms according to the accuracy metric are given in the graph.

In the classification of healthy controls, pancreatic disorders, and patients with PDAC, the most successful classification method was CV-5, while the GBC (Acc. = 72.91%) model was (AUC = 0.873). In Figure 8, the results of all classification algorithms according to the accuracy metric are given in the graph.

CA19-9 has been studied as a biomarker in many cancers since its discovery, but the highest sensitivity and specificity were obtained in pancreatic cancer patients [39, 40]. Recent studies show that the combination of C19-9 and markers can improve diagnostic accuracy [18, 41]. As seen in Figure 8 as a result of classification, the most important feature is CA 19-9, which has been studied extensively and is the gold standard [10]. The next important features are LYVE-1, TFF1, age, REG1B, creatinine, REG1A, and sex. Less than 10% of total pancreatic cancer cases occur under the age of 55. The average age of onset of pancreatic cancer is 71 [42]. Therefore, age is an important feature in classification. Studies in the literature using biomarkers obtained from various clinical data and different classification algorithms to detect pancreatic cancer were reviewed.
Detecting PDAC at an early stage for effective treatment is currently the most important challenge. In this study, the classification success of machine learning models for early-stage PDAC diagnosis with the combination of invasive urine biomarkers and CA19-9 was analyzed. It is aimed to determine the best classifying machine learning classifiers among the models used. It is a remarkable finding that ensemble learning classifiers were more successful in all our groups. The successful results of ensemble learning classifiers support the idea stated in [30] that ensemble methods such as GBC, LightGBM give superior results compared to other widely used machine learning algorithms. As a result, healthy controls, pancreatic disorders, and individuals with PDAC disease have been successfully classified. Further prospective validation studies are needed before machine learning can be applied to clinical use for PDAC diagnosis with biomarkers.
Figure 8. Accuracy of each classification algorithm (healthy controls-pancreatic disorders-PDAC patients).

Table 5. Various biomarkers, classifier models and evaluation metrics of different classification algorithms are given in the literature to detect pancreatic cancer.

<table>
<thead>
<tr>
<th>Study</th>
<th>Biomarkers</th>
<th>Best classifier models</th>
<th>Evaluation results%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honda et al. [15]</td>
<td>Plasma biomarkers</td>
<td>SVM</td>
<td>Acc. = 91</td>
</tr>
<tr>
<td>Debernardi et al. [18]</td>
<td>Urine biomarkers</td>
<td>LR</td>
<td>Control-PDAC: AUC = 99.2 Benign-PDAC: AUC = 91.9</td>
</tr>
<tr>
<td>Narayanan et al. [43]</td>
<td>Urine biomarkers</td>
<td>RF and LR</td>
<td>Prec. = 91, F1-score = 91</td>
</tr>
<tr>
<td>This study</td>
<td>Urine biomarkers</td>
<td>GBC-Light GBM</td>
<td>Control-PDAC (GBC, Acc. = 92.99 AUC = 97.61, Prec. = 93.68) PD-PDAC (Light GBM, Acc. = 86.37 AUC = 93.48, Prec. = 87.59) Control-PD-PDAC (GBC, Acc. = 72.91 AUC = 87.33, Prec. = 73.10)</td>
</tr>
</tbody>
</table>

References


