Pancreatic ductal adenocarcinoma (PDAC) is a deadly disease even in the localized stage. The 5-year survival is approximately 10% in the USA, and 80-85% of the patients present unresectable or metastatic cancer.\(^1,2\) Even for individuals eligible for surgery, the prognosis is dismal, with only 20% surviving 5 years.\(^3\) While recent advances in surgery and adjuvant chemotherapy have improved survival times, the outcomes are still far from desired. PDAC remains the fourth leading cause of cancer-related deaths in the world.\(^4\) A significant obstacle is the considerable heterogeneity of the disease and the lack of reliable clinical risk stratification. Early-stage colorectal cancer and breast cancer do not present these disadvantages, and hence, it is easy to tailor treatment selection. In the constantly changing paradigm of new treatments, it is essential to identify prediction strategies for selecting the best treatment for the right patient.\(^5\) However, current knowledge currently lacks such guiding ability.

Most recent research on localized PDAC has focused on detecting targetable genomic aberrations. However, limited success has been achieved due to the relative rarity of these alterations.\(^6\) To select the optimal treatment for each patient, the precise assumption of survival is paramount.\(^7\) However, traditional survival models such as Cox Proportional-hazards assume a linear or log-linear relationship of covariates to estimate the survival. This approach...
precludes including complex relationships between the covariates that are essential for individualized decision-making or precision medicine. Machine learning has been proposed as a feasible approach to tackle this problem. However, it has reached a performance comparable to Cox models very recently and the use of machine learning-based survival prediction in real-life datasets has been rarely studied in PDAC.\textsuperscript{8,9}

The overarching hypothesis of the current study is that machine learning models will outperform traditional survival models to predict disease-free survival (DFS) and overall survival (OS) in operated PDAC cases. This hypothesis was tested in a single-institution cohort from a tertiary reference center.

**MATERIAL AND METHODS**

**PATIENT SELECTION**

The resected PDAC cases from a university cancer center in Turkey between 2005 and 2017 were retrospectively reviewed. The study was approved by the Hacettepe University Institutional Review Board in compliance with the Helsinki Declaration (no: GO 21/489, date: 6.4.2021). The cases had a histological diagnosis of PDAC and had undergone an R0 or R1 resection. The exclusion criteria were the presence of distant metastasis at the initial presentation, R2 resection, neoadjuvant treatment, and irretrievable clinical, laboratory, or survival data. The population size was not calculated to test the maximum performance of the models with all the available cases.

**VARIABLES AND OUTCOMES**

The following data were collected from the electronic health records and patient files: demographic features, laboratory data, reports of cross-sectional imaging and pathology, operational documents, and survival times. These variables were incorporated into the machine learning models as described below. The primary outcome was the performance of the machine learning models in predicting over 6 months of DFS and over 12 months of overall survival.

**MODELS, FEATURE SELECTION, AND MODEL TRAINING**

The machine learning methods were adapted to test the performance of this approach for the problem. From our database, two machine learning models—one deep learning model, DeepHit, and the other gradient boosting decision tree model, LightGBM (Light Gradient Boosting Machine), were constructed. Technical details of these architectures are beyond the scope of this article, so they are briefly described to provide a perspective to the reader. The DeepHit model adapts the survival models with multilayered neural networks. It is a multi-task neural network, which consists of a shared sub-network and cause-specific sub-networks. DeepHit employs a residual connection (skip connection) from the input covariates into the input of each cause-specific sub-network. It is trained with a custom loss function designed to handle censored data. This loss function is the sum of 2 terms: The log-likelihood of the joint distribution of the first hitting time and event and a combination of cause-specific ranking of loss functions. LightGBM is a gradient boosting framework that uses a tree-based learning algorithm. A major difference between LightGBM and other decision tree-based algorithms lies in the construction of the trees. LightGBM does not grow a tree level-wise. Instead, it grows trees leaf-wise. It chooses the leaf it believes will yield the largest decrease in loss. LightGBM implements a highly optimized histogram-based decision tree learning algorithm, which yields great efficiency and memory consumption. The LightGBM algorithm utilizes two novel techniques called Gradient-Based One-Side Sampling and Exclusive Feature Bundling. They allow the algorithm to run faster while maintaining a high level of accuracy. LightGBM is trained with both regression loss (root-mean-square error) and classification loss (binary cross-entropy) for this problem.\textsuperscript{10,11} For the analysis, 50 train/test splits (random shuffle split with replacement) were used to estimate the generalization error. The models were trained and evaluated 50 times with different training and test sets. The number of iterations was selected as 50, and the test size was specified as 20% to obtain statistically significant results. These values work for both small and large datasets. The mean score of 50 models was the estimated generalization error, and the standard deviation was the confidence interval. The models were constructed and used in Python. The predictive ability of the models in each validation fold was assessed...
using Harrell’s concordance index, and the overall performance of the models was represented by the median score computed over 50 cross-validation cycles.

STATISTICAL ANALYSIS

Firstly, the baseline characteristics of the cases were reported with descriptive analyses of the mean (standard error) for parametric variables, median (interquartile range or range as indicated) for non-parametric variables, and frequency and percentages for categorical variables. Then, the differences between the survival groups were appropriately tested for statistical significance with the Chi-square, Mann-Whitney U, and linear regression tests. The OS time was defined as the period from surgery to the last follow-up or death. The DFS time was defined as the period between surgery to disease progression or death.

The univariate survival analysis was performed with Kaplan-Meier curves, and comparisons between the groups were performed via the log-rank test. Conventional multivariable survival analysis was performed for DFS and OS using the Cox proportional-hazards regression analysis. Lastly, the performance of the Cox models to machine learning models was evaluated from the area under the receiver operator characteristic curves (AUROC). All statistical studies were performed using the Statistical Package of Social Sciences (SPSS) Version 27.0 (Armonk, NY: IBM Corp). A type-I error level of 5% (p<0.05) was considered the threshold limit for statistical significance.

RESULTS

STUDY POPULATION AND CHARACTERISTICS

From the institutional retrospective database, 121 PDAC cases that fulfilled the inclusion criteria were included. The median age of the population was 62 (minimum-maximum: 29-88). There was male predominance (65% vs. 35%). Among the tumors, 82% were located in the pancreatic head and neck, 12% in the pancreatic tail, and 6% in the pancreatic body. The 1, 2, and 3 stages of the tumors were presented by 33, 60, and 28 patients, respectively; no patient had stage 4 cancer. All patients met the resection criteria at presentation. According to tumor localization, they underwent resection surgery with curative intent in the Whipple procedure or distal pancreatectomy. After the surgery, 85 (71%) patients had negative resection margins (R0). The remaining patients had only microscopic residual disease. Histopathologic characteristics of the tumors are presented in Table 1. Among them, 78% of patients received adjuvant chemotherapy and chemoradiotherapy in the form of Gemcitabine, Cisplatin, Taxanes, 5-FU, and their combinations.

SURVIVAL CHARACTERISTICS

During a median follow-up of 18.4 months (IQR: 9.8-36.7), 91 (75.2%), patients died and 98 (81%) patients had DFS events. The median OS of the study population was 21.9 (IQR: 11.5-44.4) months, and the median DFS was 11.8 (IQR: 6-25.6) months. The baseline patient characteristics were broadly similar in patients with disease-free survivals shorter or longer than 6 months (Table 2). Baseline characteristics were also similar in patients with overall survivals shorter or longer than 12 months, albeit vascular invasion, thrombosis, and perineural invasion were different (Table 3). In univariate analyses, the presence of tumoral thrombosis (p=0.003), pathological N stage (p<0.001), and higher postoperative carbohydrate antigen 19-9 (CA 19-9) levels (<100 vs. >100, p<0.001) were associated with impaired OS; there was a trend toward better OS in the presence of diabetes before the diagnosis (p=0.113). The DFS analyses were consistent with the OS analyses.

A multivariable analysis model was constructed using the clinical parameters with a p value of <0.15. The consistent association of perineuronal invasion and pathological T stage with OS and DFS in the literature necessitated their use as the adjustment parameters. Higher postoperative CA-19 levels [hazard ratios (HR): 2.643, 95% confidence intervals (CI): 1.288-5.426, p=0.008] and higher pathological N stage (for N1, HR: 2.248, 95% CI: 1.052-4.801, p=0.036; for N2, HR: 4.073, 95% CI: 1.725-9.617, p<0.001) were associated with lower OS. For DFS, higher postoperative CA 19-9 levels (HR: 2.472, 95% CI: 1.178-5.187) and node positivity (for N1, HR:...
1.918, 95% CI: 0.951-3.870, p=0.069; for N2, HR: 2.814, 95% CI: 1.220-6.491, p=0.015) were associated with poor outcome.

A model was constructed using the statistically significant clinical parameters (postoperative CA 19-9 levels and pathological N stage). The 0-1 system was used for the coding (for CA 19-9 levels <100, the code was 0 and for levels >100, the code was 1. For the pathological N stage N0=0, N1=1, and N2=2), so the patients had scores from 0 to 3. The AUROC analyses demonstrated AUC values of 0.752 (95% CI: 0.625-0.880, p<0.001) and 0.668 (95% CI: 0.506-0.830, p=0.045) for the 12-month OS and 6-month DFS, respectively.

MODELS

Deep learning models were constructed to predict 6 months of disease-free survival, 12 months, and 24 months of OS in the pancreatic cancer cohort. Selected features were diabetes at presentation, tumoral thrombosis, pathological T stage, pathological N stage, postoperative CA 19-9 levels, and perineural invasion. Two different machine learning models were constructed: DeepHit and LightGBM. The outputs from the DeepHit and LightGBM models for DFS and OS with the AUCs, respectively, were as follows: Relapse at 6 months was 0.58 (±0.177) and 0.73 (±0.098); survival over 12 months was 0.56 (±0.14) and 0.78 (±0.078); survival over 24 months was 0.53 (±0.13) and 0.63 (±0.083). These results are summarized in Figure 1 and Table 4. The performance of the Cox models was compared with that of the machine learning models using the AUROC (Figure 2). Across 50-repeat, 5-fold cross-validation, the median C-index (IQR) was 0.71 (IQR: 0.62-0.83) for the 12-month OS survival and 0.68 (IQR: 0.59-0.75) for the 6-month DFS (Figure 3 and Figure 4).
In this study, the performance of machine learning algorithms in predicting early relapse and medium-term mortality was tested in resectable non-metastatic PDAC and compared with the Cox proportional-hazards model. The results demonstrated that the performance of the LightGBM machine learning models was comparable to that of the Cox proportional-hazards model in this patient group. As presented above, the AUROC analyses of the LightGBM machine learning model were 0.73 (±0.098) for the 6-month DFS and 0.78 (±0.078) for the 12-month OS. In comparison, the AUC values for the Cox proportional-hazards model were 0.752 (95% CI: 0.625-0.880, p<0.001) and 0.668 (95% CI: 0.506-0.830, p=0.045) for the 12-month OS and 6-month DFS, respectively.
The performance of the LightGBM model, but not of the DeepHit model, was similar to that of the Cox proportional-hazards model. The respective AUROC values of 0.58 (±0.177), 0.56 (±0.14), and 0.53 (±0.13) for the 6-month DFS, 12-month OS, and 24-month OS of the other algorithm did not yield an acceptable performance for prediction. This observation underlined that different machine learning models could yield significantly different results depending on the dataset and the problem. To the best of our knowledge, there have been no efforts in predicting the prognosis of localized PDAC with machine learning methods.

<p>| TABLE 3: Comparison of patients with overall survivals shorter or longer than 12 months. |
|---------------------------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Longer than 12 months</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>28 (71.8%)</td>
<td>51 (62.2%)</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>65 (0%)</td>
<td>61 (0%)</td>
<td></td>
</tr>
<tr>
<td>&lt;65 years</td>
<td>18 (46.2%)</td>
<td>55 (67.1%)</td>
<td>0.028*</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td>21 (53.8%)</td>
<td>27 (32.9%)</td>
<td></td>
</tr>
<tr>
<td>TNM stage</td>
<td>Stage 1</td>
<td>5 (12.8%)</td>
<td>28 (34.1%)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>20 (51.3%)</td>
<td>40 (48.8%)</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>14 (35.9%)</td>
<td>14 (17.1%)</td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Location of tumor</td>
<td>Head</td>
<td>35 (89.7%)</td>
<td>64 (78%)</td>
</tr>
<tr>
<td>Body</td>
<td>2 (5.1%)</td>
<td>5 (6.1%)</td>
<td></td>
</tr>
<tr>
<td>Tail</td>
<td>2 (5.1%)</td>
<td>13 (15.9%)</td>
<td></td>
</tr>
<tr>
<td>Uncinate</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Weight loss at presentation</td>
<td>Absent</td>
<td>6 (15.4%)</td>
<td>14 (17.1%)</td>
</tr>
<tr>
<td>Present</td>
<td>19 (48.7%)</td>
<td>39 (47.6%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>14 (35.9%)</td>
<td>29 (35.4%)</td>
<td></td>
</tr>
<tr>
<td>Obstruction at presentation</td>
<td>Absent</td>
<td>5 (12.8%)</td>
<td>23 (28%)</td>
</tr>
<tr>
<td>Present</td>
<td>34 (87.2%)</td>
<td>59 (72%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Diabetes at presentation</td>
<td>Absent</td>
<td>25 (64.1%)</td>
<td>47 (57.3%)</td>
</tr>
<tr>
<td>Present</td>
<td>14 (35.9%)</td>
<td>35 (42.7%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Resectability at presentation</td>
<td>Resectable</td>
<td>28 (71.8%)</td>
<td>69 (84.1%)</td>
</tr>
<tr>
<td>Borderline</td>
<td>9 (23.1%)</td>
<td>11 (13.4%)</td>
<td></td>
</tr>
<tr>
<td>Unresectable</td>
<td>2 (5.1%)</td>
<td>2 (2.4%)</td>
<td></td>
</tr>
<tr>
<td>Type of surgery</td>
<td>Whipple’s</td>
<td>37 (94.9%)</td>
<td>69 (84.1%)</td>
</tr>
<tr>
<td>Distal pancreatectomy with splenectomy</td>
<td>2 (5.1%)</td>
<td>13 (15.9%)</td>
<td></td>
</tr>
<tr>
<td>Residual status</td>
<td>R0 resection</td>
<td>28 (71.8%)</td>
<td>57 (69.5%)</td>
</tr>
<tr>
<td>R1 resection</td>
<td>11 (28.2%)</td>
<td>25 (30.5%)</td>
<td></td>
</tr>
<tr>
<td>R2 resection</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Vascular invasion</td>
<td>Absent</td>
<td>30 (76.9%)</td>
<td>74 (86.2%)</td>
</tr>
<tr>
<td>Present</td>
<td>9 (23.1%)</td>
<td>8 (9.8%)</td>
<td></td>
</tr>
<tr>
<td>Thrombosis</td>
<td>Absent</td>
<td>26 (66.7%)</td>
<td>73 (89%)</td>
</tr>
<tr>
<td>Present</td>
<td>13 (33.3%)</td>
<td>9 (11%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
<td></td>
</tr>
<tr>
<td>Perineural invasion</td>
<td>Absent</td>
<td>2 (0%)</td>
<td>7 (8.5%)</td>
</tr>
<tr>
<td>Present</td>
<td>25 (64.1%)</td>
<td>60 (73.2%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>14 (35.9%)</td>
<td>15 (18.3%)</td>
<td></td>
</tr>
</tbody>
</table>

TNM: Tumor-node-metastasis.* <0.05.
Pancreatic cancer needs new strategies to combat its morbidity, mortality, and healthcare burden. Research efforts to date have yielded extensive data that include but are not limited to clinical, histopathologic, radiologic, genomic variables, and complex interactions in between these domains.12-15 This knowledge base should be used as a backbone for a precision medicine approach. However, the employment of this big data remains limited, and the adoption of machine learning can be addressed by this limitation.

**TABLE 4:** Prediction performances of the Cox and machine learning models.

<table>
<thead>
<tr>
<th></th>
<th>6 month disease-free survival (AUROC±SD)</th>
<th>12-month overall survival (AUROC±SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox proportional hazards model</td>
<td>0.35*</td>
<td>0.26**</td>
</tr>
<tr>
<td>DeepHit</td>
<td>0.58 (±0.177)</td>
<td>0.56 (±0.14)</td>
</tr>
<tr>
<td>LightGBM</td>
<td>0.73 (±0.098)</td>
<td>0.78 (±0.078)</td>
</tr>
</tbody>
</table>

*p<0.045; **p<0.001; AUROC: Area under the receiver operator characteristic curves; SD: Standard deviation.

**FIGURE 1:** Kaplan-Meier curve of the study population.

**FIGURE 2:** Outputs of machine learning models; the line represents the mean value of 50 different models with each approach. A: DeepHit models for the 6-month disease-free survival. B: Light Gradient Boosting Machine (LightGBM) models for the 6-month disease-free survival. C: DeepHit models for the 6-month overall survival. CD: Light Gradient Boosting Machine models for 12-month overall survival. Disease-free and overall survivals for DeepHit are as follows: Relapse at 6 months, 0.58 (±0.177); survival over 12 months, 0.56 (±0.14); survival over 24 months, 0.53 (±0.13). Disease-free and overall survivals for LightGBM are as follows: Relapse at 6 months, 0.73 (±0.098); survival over 12 months, 0.78 (±0.078); survival over 24 months, 0.63 (±0.083).
Artificial intelligence methods have been tested to tackle other problems with pancreatic cancer.16-20 Most studies focused on the early detection and screening of pancreatic cancer, working on strategies that utilized imaging, blood-based, and clinical data. Only 20-30% of the patients are currently diagnosed while the disease is localized.3 If possible, adopting these efforts in clinical practice will enable earlier treatment, and thus, more prolonged survival. However, we will need additional strategies for optimal treatment of localized PDAC as their 5-year survivals are around 40%.21,22

The prognosis of resectable pancreatic cancer has been studied extensively. Currently, known clinical predictors of survival are age, gender, neoadjuvant treatment, CA 19-9 levels, diabetes at presentation, tumor-node-metastasis (TNM) stage, tumor size, tumor grade, tumor location. Nomograms including these clinical parameters have also been used to precisely predict patient outcomes.2,23,24 Other prognostic biomarkers proposed for pancreatic cancer include histopathologic markers such as tumor budding, vascular/perineural invasion, desmoplastic reaction, etc. Molecular markers such as DKC, MUC4, hENT1, micro-RNA profiles, circulating tumor DNA, and circulating tumor cells can also be utilized.25 While the standard clinical criteria were used in the prognosis prediction models, additional adjustments according to molecular markers could not be made due to a lack of data.

Machine learning models have been studied in breast, lung, central nervous system, urogenital, and gastrointestinal cancers. An extensive review is beyond the scope of the current discussion.26-30 Briefly, these studies utilized national and institutional databases and multi-omic data and mostly yielded better predictions compared to Cox and linear regression.31 In the current study, one of the machine learning
models and the Cox-regression-based dichotomous model displayed a similar performance, possibly due to the small sample size.

The current study had several limitations inherent to its small population size, retrospective design, and machine learning methods. A localized and regional pancreatic cancer population was included that had undergone curative surgeries and continued management in a single institution. Additionally, most patients were not treated with the current standards of care adjuvant regimens, limiting result generalization to current practice. This population was selected because the study was believed to have maximum impact on their management. However, the drawbacks of small population size, such as generalizability and statistical power, are acknowledged. Machine learning has its limitations, such as over-fitting the model to the dataset; although, cross-validations were performed. Further validation of this model in a different and larger dataset is required.

## CONCLUSION

In conclusion, machine learning should be adopted to a problem in medicine that connects any data virtually to a measurable outcome, including the prognosis prediction in difficult-to-treat cancers like pancreatic cancer. The future of machine learning in this field depends on further efforts with more extensive and comprehensive datasets followed by prospective studies.

## Source of Finance

This study is supported by Algedicus Artificial Intelligence and Medical Simulation Company, Proje no: (21-003). Ankara, Turkey.

## Conflict of Interest

Cem Simsek is a shareholder in Algedicus Artificial Intelligence and Medical Simulation Company.

## Authorship Contributions

**Idea/Concept:** Cem Şimşek, Deniz Can Güven, Şuyib Yalçın; **Design:** Cem Şimşek, Şuyib Yalçın; **Control/Supervision:** Cem Şimşek, Yasemin Balaban, Şuyib Yalçın; **Data Collection and/or Processing:** Cem Şimşek, Furkan Ceylan, İbrahim Yavva Çakır, Taha Koray Şahin; **Analysis and/or Interpretation:** Cem Şimşek, Deniz Can Güven, Ömer Dizdar; **Literature Review:** Cem Şimşek, Deniz Can Güven, Şuyib Yalçın; **Writing the Article:** Cem Şimşek, Deniz Can Güven, Şuyib Yalçın; **Critical Review:** Cem Şimşek, Deniz Can Güven, Taha Koray Şahin; **References and Fundings:** Cem Şimşek, Şuyib Yalçın; **Materials:** Cem Şimşek.

## REFERENCES

5. Herreros-Villanueva M, Bujanda L. Non-invasive biomarkers in pancreatic cancer diagnosis: What we need versus what we have. Ann Transl Med. 2016;4(7):134. [Crossref] [PubMed] [PMC]
11. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;785-794. [Crossref]


27. Xie Y, Meng WY, Li RZ, et al. Early lung cancer diagnostic biomarker discovery by machine learning methods. Transl Oncol. 2021;14(1):100907. [Crossref] [Pubmed] [PMC]


