Artificial neural networks predict survival from pancreatic cancer after radical surgery

Daniel Ansari, M.D., b, Johan Nilsson, M.D., Ph.D., a, Roland Andersson, M.D., Ph.D., a, Sara Regnér, M.D., Ph.D., c, Bobby Tingstedt, M.D., Ph.D., a, Bodil Andersson, M.D., Ph.D., a,*

aDepartment of Surgery, Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden; bDepartment of Cardiothoracic Surgery, Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden; cDepartment of Surgery, Clinical Sciences Malmö, Lund University, Skåne University Hospital, Malmö, Sweden

KEYWORDS: Pancreatic cancer; Artificial neural network; Surgery; Survival; Prognosis

Abstract

BACKGROUND: Artificial neural networks (ANNs) are nonlinear pattern recognition techniques that can be used as a tool in medical decision making. The objective of this study was to develop an ANN model for predicting survival in patients with pancreatic ductal adenocarcinoma (PDAC).

METHODS: A flexible nonlinear survival model based on ANNs was designed by using clinical and histopathological data from 84 patients who underwent resection for PDAC.

RESULTS: Seven of 33 potential risk variables were selected to construct the ANN, including lymph node metastasis, differentiation, body mass index, age, resection margin status, peritumoral inflammation, and American Society of Anesthesiologists grade. Three variables (ie, lymph node metastasis, leukocyte count, and tumor location) were significant according to Cox regression analysis. Harrell’s concordance index for the ANN model was .79, and for Cox regression it was .67.

CONCLUSIONS: For the first time, ANNs have been used to successfully predict individual long-term survival for patients after radical surgery for PDAC.

Pancreatic cancer is the second most frequent gastrointestinal malignancy and the fourth leading cause of cancer death in Western societies. Adenocarcinoma derived from ductal cells of the exocrine pancreas constitutes the majority of all pancreatic cancers. Prognosis remains poor with less than 20% having resectable disease and a collective 5-year survival rate of at most 6%.

Outcome prediction is important in the clinical decision-making process. Currently, the tumor node metastasis (TNM) classification from the American Joint Committee on Cancer is considered the gold standard to estimate prognosis in pancreatic ductal adenocarcinoma (PDAC) and other malignancies. However, heterogeneity in tumor and patient characteristics makes this system rather nondiscriminatory. For example, microscopic tumor clearance, histologic grade, and comorbidity may also influence survival after resection with curative intent. Thus, the addition of other variables in addition to the TNM stage may improve the prediction of outcomes in patients with PDAC and may be used for patient counseling, treatment selection, and clinical trial design.

Artificial neural networks (ANNs) work in a nonlinear fashion, which may better describe the interactions between health risk factors than traditional statistical models. ANNs are increasingly being used in complex medical decision making and have been used to predict outcome in patients with various malignancies, including breast.
colonic, and esophageal cancer. The Cox proportional hazards regression is a commonly used statistical technique for survival analysis. To use the ANN in long-term survival prediction, a different approach is necessary. Such a model has been described by Biganzoli et al.

The aim of the present study was to construct and validate an ANN model for predicting long-term survival in patients with PDAC who have undergone surgical resection with curative intent by using data from a single institution. The predictive ability of the ANN was compared with Cox regression.

Methods

Study population

All consecutive patients undergoing macroscopically radical resection for histologically proven PDAC at the Department of Surgery, Skåne University Hospital, Lund and Malmö, Sweden, between January 1995 and September 2010 were identified from the hospital records aided by a computer search (International Classification of Diseases, Ninth Revision code C25.0-C25.9). The case records were retrospectively reviewed.

Thirty-three input variables were collected and considered appropriate for analysis as potential risk variables. These included demographics (ie, age and sex), clinical factors (ie, body mass index [BMI], American Society of Anesthesiologists [ASA] grade, diabetes mellitus, smoking, jaundice, biliary drainage, and double duct sign), standard preoperative laboratory tests (ie, white blood cell count, hemoglobin, platelet count, aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, γ-glutamyl transpeptidase, and total bilirubin), operative details (ie, length of stay and adjuvant therapy).

Study definitions

Smoking was defined by a positive history of smoking within 5 years of cancer diagnosis. Jaundice was characterized by a total serum bilirubin concentration of 3 mg/dL (≈50 μmol/L) or higher. The definition of a double duct sign was the simultaneous dilation of the common bile duct and pancreatic duct, with biliary strictures in the head of the gland. Transfusion was defined as any autologous or allogeneic red blood cell transfusion received from the time of operation until discharge. Resection margins were categorized as R0 (ie, no residual tumor) or R1 (ie, microscopic residual tumor). A margin was designated R1 if tumor cells were present at the surface of the resection margin. Peritumoral inflammation was defined as a chronic inflammatory infiltrate within the pancreatic parenchyma. It was graded as (−) when inflammatory reaction was sparse or absent and (+) when moderate or severe inflammation was described.

Calibration and validation of the ANN model

A feed-forward ANN was used to construct the survival model. Several ANNs were combined into a single prediction model (ie, a committee machine). To provide an estimation of the conditional probability of the event in question as a function of both risk variables and time, the model described by Biganzoli et al as a generalization of the standard Cox proportional hazard model was used. Each multilayer perceptron was trained using conjugate gradient descent applied to an entropy error function. The number of hidden nodes and members in the committee machine were determined based on experiments, starting with a single node and increasing the number of nodes until the highest performance was found using Harrell concordance index (C-index).

To avoid overtraining and improve the generalization performance, a weight decay term was used. The calibration of the model was performed using a 10-fold cross-validation procedure. To select the most important risk variables and to minimize the number of variables included in the final model, a ranking of risk variables was performed. A baseline C-index was calculated using all variables. The order of relevance was obtained by measuring the change in index when 1 risk variable was omitted from the model. This procedure was repeated for each of the variables included, and the least relevant variable was omitted from the model. To optimize the model, the bottomed ranked variable was eliminated, the ANNs were recalibrated using n-1 variables, and a new identification procedure of the least relevant variable was performed. This procedure was repeated until only 1 variable remained. The final ranking list was constructed from the top-ranked variables, which improved the performance of the model.

Statistical analyses

Continuous variables are presented as median (interquartile range) and categoric variables as the total number or frequencies. Multivariate analysis by Cox proportional hazard model was performed using stepwise Cox regression. Inclusion criterion for the full model was $P < .250$. The limit for stepwise backward elimination was $P < .100$. The Kaplan-Meier method was used to estimate long-term survival. The performance of the survival model was determined using C-index. Missing data were handled using multiple imputation techniques. High-performance computing
clusters were used to train and evaluate the ANNs. The ANN calibration and analyses were performed with MatLab Distribution Computing Server 2010a, Neural Network Toolbox (MathWorks, Natick, MA). Statistical analyses were performed using the Stata MP version 11.1 (2010) statistical package (StataCorp LP, College Station, TX).

Results

During the study period, 84 patients underwent pancreatic resection for PDAC and were included in the study (Table 1). The median age was 66 years (58–72 years), and 48 of patients were men (57%). A pancreaticoduodenectomy was performed in 77 patients (92%) and a distal pancreatectomy in 7 patients (8%). Eleven patients (13%) developed major postoperative complications, including anastomotic leak in 4 patients, wound dehiscence in 2 patients, myocardial infarction in 1 patient, infected seroma in 1 patient, pancreatic fistula in 1 patient, intra-abdominal hematoma in 1 patient, and mucosal bridge in 1 patient. Forty-five patients received adjuvant gemcitabine, capecitabine, or 5-flourouracil–based chemotherapy after surgical resection. The median survival of patients who received adjuvant therapy was 28 months compared with 18 months for observation. The median survival of the entire cohort was 19 months.

Table 1  Patient characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Median (IQR) or n (%)</th>
<th>% Missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>66 (58–72)</td>
<td>0</td>
</tr>
<tr>
<td>Sex ratio (M:F)</td>
<td>48:36</td>
<td>0</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25 (22–27)</td>
<td>10</td>
</tr>
<tr>
<td>ASA grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>15 (18)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>46 (55)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>23 (27)</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>23 (27)</td>
<td>0</td>
</tr>
<tr>
<td>Smoking</td>
<td>33 (43)</td>
<td>10</td>
</tr>
<tr>
<td>Jaundice</td>
<td>67 (80)</td>
<td>0</td>
</tr>
<tr>
<td>Preoperative biliary drainage</td>
<td>57 (70)</td>
<td>0</td>
</tr>
<tr>
<td>Double duct sign</td>
<td>57 (76)</td>
<td>11</td>
</tr>
<tr>
<td>WBC count (×10⁹/L)</td>
<td>7.2 (6.4–8.5)</td>
<td>6</td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>133 (122–142)</td>
<td>0</td>
</tr>
<tr>
<td>Platelet count (×10⁹/L)</td>
<td>245 (203–298)</td>
<td>1</td>
</tr>
<tr>
<td>AST (μkat/L)</td>
<td>.7 (.4–1.4)</td>
<td>0</td>
</tr>
<tr>
<td>ALT (μkat/L)</td>
<td>.9 (.5–2.5)</td>
<td>0</td>
</tr>
<tr>
<td>ALP (μkat/L)</td>
<td>3.9 (1.9–11)</td>
<td>2</td>
</tr>
<tr>
<td>GGT (μkat/L)</td>
<td>2.2 (.9–9.0)</td>
<td>12</td>
</tr>
<tr>
<td>Total bilirubin (μmol/L)</td>
<td>25 (11–97)</td>
<td>0</td>
</tr>
<tr>
<td>Estimated blood loss (mL)</td>
<td>1,000 (700–1,400)</td>
<td>0</td>
</tr>
<tr>
<td>Blood transfusion</td>
<td>51 (61)</td>
<td>0</td>
</tr>
<tr>
<td>Tumor location (head)</td>
<td>77 (92)</td>
<td>0</td>
</tr>
<tr>
<td>Tumor size</td>
<td>3.0 (2.0–3.5)</td>
<td>0</td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>7 (8)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>24 (29)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>52 (63)</td>
<td></td>
</tr>
<tr>
<td>Positive surgical margin</td>
<td>31 (37)</td>
<td>0</td>
</tr>
<tr>
<td>Tumor differentiation (poor)</td>
<td>38 (48)</td>
<td>6</td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td>47 (57)</td>
<td>1</td>
</tr>
<tr>
<td>No. of positive nodes</td>
<td>1 (0–2)</td>
<td>8</td>
</tr>
<tr>
<td>Vascular invasion</td>
<td>11 (13)</td>
<td>1</td>
</tr>
<tr>
<td>Perineural invasion</td>
<td>40 (48)</td>
<td>1</td>
</tr>
<tr>
<td>Adipose tissue invasion</td>
<td>36 (43)</td>
<td>1</td>
</tr>
<tr>
<td>Peritumoral inflammation</td>
<td>22 (27)</td>
<td>1</td>
</tr>
<tr>
<td>Clavien grade ≥ 3</td>
<td>11 (13)</td>
<td>0</td>
</tr>
<tr>
<td>Length of stay (d)</td>
<td>16 (13–21)</td>
<td>0</td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td>45 (54)</td>
<td>0</td>
</tr>
</tbody>
</table>

ALP = alkaline phosphatase; ALT = alanine aminotransferase; ASA = American Society of Anesthesiologists; AST = aspartate aminotransferase; BMI = body mass index; F = female; GGT = γ-glutamyl transpeptidase; M = male; WBC = white blood cell.

ANN architecture

Six million different ANN models were validated using a high-performance computer cluster. The architecture of the final ANN consisted of 4 hidden nodes and 1 output node (Fig. 1). This ANN architecture was used to identify the most influential risk variables that can be associated with survival.

Performance and accuracy

Fig. 2 shows the change in the performance of the ANN model with the number of input variables plotted on the
x-axis and C-index on the y-axis. The first peak was achieved when the 7 top-ranked variables were selected. Included herein were lymph node metastasis, poor differentiation, BMI, age, positive resection margin, peritumoral inflammation, and ASA grade (Table 2). The Cox regression model selected 3 variables: lymph node metastasis, tumor location, and preoperative white blood cell count. The ANN model was more accurate than Cox regression in predicting survival. The C-index was .79 for the ANN and .67 for Cox regression.

Time dependence of hazard ratios

In Cox regression, it is assumed that the hazard ratio proportion remains constant throughout the time period of the analysis. However, a more accurate representation of time is to include it as a covariate. This approach is shown in Fig. 3. For example, the hazard ratio for lymph node metastasis was relatively stable over time. In contrast, the hazard for ASA grade gradually increased.

Individualized survival prediction

The ANN survival predictions were compared with observed survival using the Kaplan-Meier method. Fig. 4 shows that ANN and Kaplan-Meier survival curves were similar, which indicated that predicted survival did not considerably differ from observed survival. This finding was further supported by the C-index. Over time, the agreement between the ANN and Kaplan-Meier decreased because the analysis was affected by the limited number of patients with actual long-term follow-up.

Comments

This is the first study showing the application of ANNs for predicting survival in pancreatic cancer. The C-index was .79, indicating a good discriminatory power. Although the current TNM staging system continues to be the standard determinant of prognosis after resection for pancreatic cancer, the heterogeneity of tumor biology and patient characteristics results in significant variation of outcome within each staging category. Thus, incorporating other important prognostic variables in addition to the TNM classification might improve the prediction of outcomes.

Several studies have investigated ANNs in the prediction of survival from other cancer types.7–9 Many of these used receiver operating characteristic curves to describe the accuracy and the discrimination for the different models. We used Harrell C-index, which is similar to receiver operating characteristic analysis in the logistic model but appropriate for censored data. The C-index provides the probability that given 2 patients, one who will die before the other, the model will assign a higher probability of death to the former.23 The C-index of previous ANN models for survival prediction in cancer has ranged from .66 to .81.24–28 Thus, in comparison with these studies, our ANN model had good predictive ability.

Based on a Cox regression model, the Memorial Sloan-Kettering Cancer Center developed a nomogram that estimates the probability that patients undergoing resection for PDAC would be alive at 1, 2, and 3 years postoperatively.29 The variables required for the Memorial Sloan-Kettering Cancer Center nomogram were age, sex, weight loss, portal vein resection, splenectomy, resection margin, tumor location, histologic differentiation, posterior resection margin, the number of positive nodes, the number of negative nodes, back pain, tumor stage, and maximal neoplasm size. This nomogram was found to be accurate by internal validation, and predictions discriminated better than the TNM system did with the C-index equal to .64 and .56, respectively. Botis et al10 developed a multivariable Cox model for predicting survival in resectable pancreatic cancer taking age, tumor differentiation, size, alkaline phosphatase, albumin, and carbohydrate antigen 19-9 (CA 19-9) into account. The C-index was .73 compared with .59 for the TNM system.

Although the Cox model is well accepted for survival analyses, it suffers from a number of limitations. For example, the impact of any prognostic marker on the analysis has to be assessed a priori, and results cannot
always be applied to individual cases.\textsuperscript{31} It also assumes that there are linear correlations and that hazards are proportional during follow-up. The ANN models may overcome the limitations of the Cox models, which may result in improved prediction.

Consistent with previous reports,\textsuperscript{3} we confirmed that lymph node metastasis, tumor differentiation, and margin status were correlated with outcome for patients undergoing resection for pancreatic cancer. BMI was also considered an important prognostic factor in the ANN. Recently, the beneficial impact of high BMI on the survival of patients undergoing pancreaticoduodenectomy for pancreatic cancer has been reported.\textsuperscript{32} The median BMI at operation was 24.5 in the present study, with 39.5% being obese or overweight. The patients with a high BMI may have more physiologic reserve than lower BMI patients and may tolerate better the effects of surgical resection and other therapies.\textsuperscript{33} In the present series, young patients had a tendency to have a worse prognosis than older patients. This has not been previously documented in the context of PDAC. However, in other types of cancer, clinical and biomarker data suggest that early-onset cancers may grow faster and be biologically more aggressive than late-onset cancers.\textsuperscript{34,35} Several mechanisms have been proposed concerning the role of inflammation in the development and progression of pancreatic cancer.\textsuperscript{36}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{Survival curves and time-dependent hazard ratios for the 7 top-ranked variables. For continuous variables (BMI and age), the time-dependent hazard ratio was calculated using the interquartile range instead of a unit difference.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{A comparison of ANN survival prediction and Kaplan-Meier estimates.}
\end{figure}
presence of peritumoral inflammation was associated with a decrease in survival in our series. Other authors have also noted the prognostic value of histologic inflammation by using the ratio of Th2/Th1 tumor-infiltrating lymphoid cells. Concerning the influence of preoperative comorbid conditions, we found that the ASA grade had a significant impact on the performance of the ANN model.

The survival of patients with PDAC is among the shortest of all cancer types. This is often attributed to a delayed diagnosis and resistance to conventional chemotherapy. ANNs can aid in the early diagnosis of PDAC using imaging findings, histologic sections, or plasma proteomic profiles. Furthermore, novel biomarkers are increasingly being discovered. Such molecular variables may be included in the ANN to improve its predictive ability.

There are several potential limitations in the present study. All the data were analyzed retrospectively and over a relatively long timeframe, with potential changes in imaging and histopathological and treatment approaches over time. The adjuvant chemotherapy regimens were heterogeneous during the time period. There may also have been a selection bias as to which patients received adjuvant treatment. This may explain why adjuvant chemotherapy was not included in either the ANN or the Cox regression model.

The percentage of missing data per variable ranged from 0% to 12% in this study. We used multiple imputation techniques to substitute missing values. Data imputation has been shown to be superior to complete case analysis and the missing indicator method in multivariable models. ANNs have suffered difficulties with generalization, producing models that can overfit (“overtrain”) the data. To improve the generalization, an internal cross-validation was performed; however, external validation would have been beneficial.

In summary, this pilot study showed that ANNs can be used to accurately predict the survival of patients after surgical resection. Further studies using large, prospective cohorts are warranted. In the future, biomarkers such as genes, microRNAs, or proteins can be incorporated into the ANN in order to increase its predictive ability.

References


