

Predictive Modelling For Early Diagnosis of Pancreatic Cancer

Arpit Tyagi
UCMS,DU
DELHI,INDIA

Sarthak Deshwal
NSIT,DU
DELHI,INDIA

Sumit
NSIT,DU
DELHI, INDIA

Rahul Yadav
NSIT, DU
DELHI, INDIA

ABSTRACT

Pancreatic carcinoma is one of the most aggressive tumors. The mortality rate is exceptionally high. There is difficulty in early diagnosis of the pancreatic carcinoma due to limit of imaging techniques and lack of precise knowledge of role of cancer biomarkers. To find a relation between cancer biomarkers level, CA 19-9, CA 125 and Pancreatic carcinoma, we can use artificial neural networks like back propagation as described in the paper which possess a distinct advantage of predicting the pancreatic carcinoma on basis of the level of these biomarkers.

Keywords: *Pancreatic carcinoma, Cancer biomarkers, CA 19-9, CA 125, Back propagation, Mortality rate*

I. INTRODUCTION

Most of the patients with pancreatic carcinoma present when the pancreatic carcinoma is locally advanced or metastasis disease in asymptomatic phase. Around 80% of the patients present late with metastasis at diagnosis^[1]. Metastasis is the most common cause of death in pancreatic carcinoma patients.

Challenges in early diagnosis of pancreatic carcinoma:

1. Patients present with non specific symptoms making diagnosis more difficult.
2. Methods used like imaging techniques are not conclusive and often of ambiguous relevance.

3. Cytological examination of samples obtained by endoscopic ultrasound-guided fine needle aspiration may not be precise due to various factors like sampling difficulties other associated pathologies^[2].

The above mentioned reasons led to search for relevant biomarkers for diagnosis of pancreatic carcinoma by biomarkers.

DIAGNOSIS USING BIO-MARKERS

CA 19-9 is the most widely used biomarker for management of pancreatic carcinoma, but due to limitations it's role is limited and used mainly for detection of tumor recurrence after surgical resection^[3]. But the use of CA 19-9 in combination with other biomarkers has proved to improve accuracy^[4]. CA 125 biomarker in combination with CA 19-9 is found to increase the sensitivity by 6% compared to the CA 19-9 essay alone. CA 125 is also of limited diagnostic value in pancreatic cancer if used alone^[5]. We can use predictive neural network algorithms to find a pattern between cancer biomarker level, CA 19-9 and CA 125 and pancreatic carcinoma^[6].

II. METHODOLOGY

1. Data collection

We have collected the data from research.fhcr.org (Pancreatic cancer diagnostic marker data). The data is a three variable type data where the first two variables are of float type (y1,y2) and the third variable is of binary type. The variable y1 and y2 are biomarkers which represent the extend of presence

of cancer in the body. The third variable d which is a binary type data indicate whether the person from the current observation has pancreatic cancer or not (1/0). We have taken 141 instances based on 2 different biomarkers level. pancreatic cancer is the least occurring and most fatal cancer so the amount of observation for this type of cancer is very less. Out of the 141 observations present at time during testing phase of the model randomly 51 observations were treated to be the test data. The class value was reported as yes (90 cases) or no (51 cases) to represent pancreatic carcinoma.

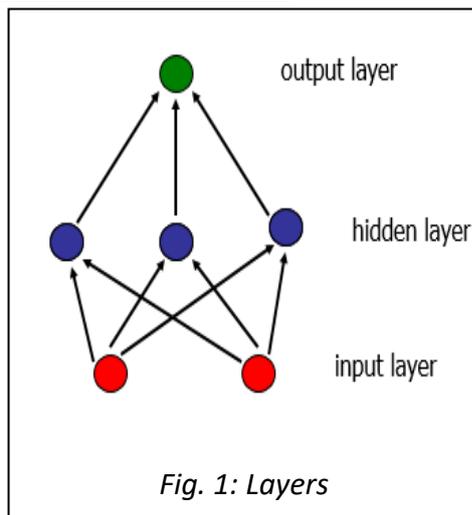
2. Data predicting algorithm

Data predicting algorithm has been used to predict the pancreatic cancer. Back propagation algorithm: Back propagation algorithm is based on the changes in the cost function due to changes in weights and biases in a network.

Back propagation

Back propagation is a training method in artificial neural networks and it is commonly used along with an optimization method such as gradient method. It uses a two phase cycle. One is propagation and another is weight update. Input vector is propagated through the network, layer by layer. Received vector is compared with the desired output and an error value is calculated for each layer.

Back propagation Algorithm



Lets us assume we have training examples as a pair of the form $e \langle (x_1, \dots, x_n), t \rangle$. Here (x_1, \dots, x_n) is the vector of input values and t is the target output value. " η " is the learning rate.

- Initialize weights for every neuron (typically random).
- Keep doing above steps until error obtained stops improving.
 - ❖ For each training example $e \langle (x_1, \dots, x_n), t \rangle$:
 - ForwardPass to compute: $O_m = \text{Net_output}(\text{Network}, e)$.
 - For each output unit m , compute: $\delta_m = O_m(1 - O_m)(t_m - O_m)$
 - For every hidden unit, k : $\delta_k = O_k(1 - O_k) \sum_m w_{k,m} \delta_m$
 - For every neuron weight W_j : $w_{i,j} = w_{i,j} + \Delta w_{i,j}$ where $\Delta w_{i,j} = \eta \delta_j x_{i,j}$

Training Data Collection

- For Online Learning: Online Learning uses dynamic environments. It helps to provide a flow of new training data patterns.
- For Offline Learning: It makes use of static data patterns.

Applications

- It is useful in learning complex mappings from input to output.
- Doesn't require high level of understanding from the user. User can be guided by heuristics.

Limitations

- Back propagation uses a gradient descent algorithm which is very slow. It uses very small learning rates for stable learning.
- For a non-linear network picking the learning rate is itself a challenge.
- Normalization of input vectors could improve performance but it is not there in back propagation.

III. RESULTS AND DISCUSSION

1. Dataset

The data set comprised of 2 attributes of 141 patients. The patients were described as either '0' or '1' on the basis of presence or absence of pancreatic cancer.

Description of the dataset is as follows:

Sl. No	Attribute name	Attribute type	Attribute description
1.	CA 19-9	Numeric	Amount of biomarker in patient
2.	CA 125	Numeric	Amount of biomarker in patient
3.	Class	Numeric [0,1]	Status of pancreatic cancer in patient

Table 1. Indicators

2. Application of predicting algorithm:

Back propagation algorithm was implemented to predict pancreatic cancer in patients of the dataset. The model was generated for training set and used for training set. It was observed that using back propagation algorithm an accuracy of 98.82% was achieved in prediction of pancreatic cancer.

IV. USES

The predictive algorithm based systems can be used to predict the possibility of pancreatic carcinoma in patients. Early diagnosis of pancreatic carcinoma will lead to decrease in mortality and better quality of life. With use of biomarkers and combination of more and more biomarker better specificity and sensitivity can be achieved. These methods being comparatively cheaper and non invasive can be really useful.

V. CONCLUSION

Identification of Pancreatic carcinoma at an early stage is important to prevent and manage mortality. The patients must be screened at an early stage based on initial symptoms to find the possibility of pancreatic carcinoma.

VI. REFERENCES

- [1]. Rosen, Harald. "Pancreatic cancer—a diagnostic and therapeutic challenge." *HAMDAN MEDICAL JOURNAL* 8.2 (2015).
- [2]. Canto, Marcia Irene, et al. "International Cancer of the Pancreas Screening (CAPS) Consortium summit on the management of patients with increased risk for familial pancreatic cancer." *Gut* 62.3 (2013): 339-347.
- [3]. Ballehaninna, Umashankar K., and Ronald S. Chamberlain. "The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal." *Journal of gastrointestinal oncology* 3.2 (2011): 105-119.
- [4]. Ferri, María José, et al. "Improved pancreatic adenocarcinoma diagnosis in jaundiced and non-jaundiced pancreatic adenocarcinoma patients through the combination of routine clinical markers associated to pancreatic adenocarcinoma pathophysiology." *PLoS one* 11.1 (2016): e0147214.
- [5]. Haglund, C. "Tumour marker antigen CA125 in pancreatic cancer: a comparison with CA19-9 and CEA." *British journal of cancer* 54.6 (1986): 897.
- [6]. Møller, Martin Fodsløtte. "A scaled conjugate gradient algorithm for fast supervised learning." *Neural networks* 6.4 (1993): 525-533