Kaplan-Meier plotter data analysis model in early prognosis of pancreatic cancer

S Zwyea*, L Naji1 and S Almansouri2

1Department of Radiological Techniques, College of Health and Medical Technology/Baghdad, Middle Technical University, Baghdad- Iraq
2Al-Nahrain University, College of Engineering, Department of Computer Engineering, Baghdad- Iraq

*Corresponding Dr.susanzwyea@mtu.edu.iq

Abstract. Intelligent data analysis methods provide helpful tools for cancer researchers to detect the prognosis of patients with specific diseases. Yet, very little information is known about the features of these models used in data analysis methods. In this study, we presented a new Kaplan-Meier plotter model with a better-combination of input features for early prognosis tasks of pancreatic cancer. Our new model integrates gender, race, and follow up the threshold to get better verification of genes of interest as prognostic markers for predicting cancer at early stages. Assessment is made for the developed model to examine the important role of the oncogene Rab1A in early prediction of pancreatic cancer on the standard clinical datasets from The Human Protein Atlas. Our results showed that overexpression of the oncogene Rab1A in pancreatic cancer plays a vital role in its early prognosis (p<0.05). The proposed model results were also verified using an independent dataset deposited in The Human Protein Atlas. Altogether, the experimental results highlight Rab1A potential role in cancer prognosis.

Keywords: Pancreatic cancer, Rab1A, Biomarker, Kaplan-Meier plotter.

1. Introduction
Pancreatic cancer has the fourth highest mortality rate among other deadly cancers [1]. The incidence of pancreatic cancer is 56770 newly diagnosed patients and 45750 deaths yearly in the United States [2]. The Pancreatic cancer survival rate is low (2-9%), despite the improvement in pancreatic cancer treatments. However, the pancreatic cancer survival rate can be improved when it is detected in the early stages. Therefore, the early the early prognosis of pancreatic tumors is essential for successful treatment. In the last decade, many researchers improve the performance of many techniques used in clinical research to ensure an early prognosis of a cancer kind. These techniques are clinical images analysis, machine learning, and miRNA expression analysis [3-9].

One of the intelligent data analysis methods is machine learning technology. Machine learning uses algorithms to analyze complex medical datasets. It has several applications in different areas in medicine such as medical prognosis and research. In recent years, machine learning shows significant improvement in predictive pancreatic cancer compared with other techniques that dominate the medical literature [10-16]. Regarding good performance and accuracy of machine learning in medical fields, it has not been accepted in practice for many reasons. Among these reasons perhaps due to increase in the abundance of tools available to physicians which further increases the complexity of the physician work. Scientists start to identify and predict diseases using clinical image analysis magnified images of lesion regions. Recently, the clinical analysis methods exhibit higher accuracy in cancer prediction in comparison with the unaided eye [17]. Due to several limitations and artifacts that are associated with clinical images which cause low accuracy, scientists tend to use other automated methods to analyze patient data. Therefore, some scientists start to combine gene expression and machine learning to predict the prognosis of patients in techniques called mRNA profiling. mRNA profiling uses mRNA to diagnose and help in predicting the outcomes of cancer patients. It also could be used as a biomarker or therapeutic target.
However, a suitable level of validation for these techniques is required to be considered in practice in medical diagnosis. Here in this work, we improve a model for pancreatic cancer prediction. in order to fulfill this goal, the presented model is assessed with a dataset placed in The Cancer Genome Research Network (TCGA) database shown in Table 1. The contributions of this work can be outlined as follows:

- The presented analysis model was assessed with pancreatic tumors RNA-seq datasets.
- The proposed model was investigated for the impact of a new set of extended features in pancreatic cancer prediction (prognosis).
- The performance of the model was assessed from TCGA datasets to The Human Protein Atlas.

| Method                          | Cancer type | Features as inputs                                      | Outputs                                           |
|---------------------------------|-------------|--------------------------------------------------------|---------------------------------------------------|
| Classical Kaplan Meier analysis model | Pancreatic  | Overall survival, Rab1A gene, 240 months threshold follow up | Prognosis of pancreatic cancer                    |
| Developed Kaplan-Meier analysis model | Pancreatic  | overall survival, Rab1A gene, 120 months threshold follow up, gender, race | Validation of Rab1A gene as a prognostic marker for pancreatic cancer in white female stage 2 |

**2. Preliminaries**

2.1 New Kaplan-Meier analysis model for pancreatic cancer early prediction

Kaplan Meier method is one of the best estimates which was designed to detect such changes for the evaluation of treatments and the prediction of their effects on cancer survival probability [18]. Kaplan Meier's classical model consists of fifteen features. These features are: gene symbol or multiple genes symbols, cutoff values, survival used to compute median survival (include relapse-free survival and overall survival) when the cohorts reach median survival, follow up threshold (referred to the period time, in months, after patients receiving treatment), quality control (removed redundant sample, check the proportional hazards assumption, and perform multivariate analysis), cancer stage (range from one to four), cancer grade (range from one to four), race (white and African American), gender, analysis based on cellular contents and cancer type (Basophils, B cells, T cells, Eosinophils, and Macrophages).

However, the classical Kaplan Meier model can only detect diseases in later stages. In most diseases especially, cancer if it is not detected and treated at its early stages, then patients will not get a full recovery of the disease. In order to improve Kaplan Meier model detection tasks for a disease type, we extended the model features by integrated new restriction features along with the basic features. The extension of the model features makes the analysis more restricted to threshold, race, and gender. The model's new set of features is important in enhancing the model detection output. The proposed model was evaluated with a dataset placed in the TCGA portal of 176 cases.

2.2 Assessment of Kaplan-Meier new model results to Human Protein Atlas databases

Pancreatic cancer is the fourth leading cause of cancer death worldwide. Detecting pancreatic cancer in its early stages plays an important role in improving cancer treatment. Finding a new prognostic biomarker for pancreatic cancer is a crucial and challenging task. Therefore, we evaluated the results of the new Kaplan Meier model on the standard clinical datasets from Human Protein Atlas (https://www.proteinatlas.org). The data was downloaded from the following website (proteinatlas.xml.gz). The downloaded data was used to analyze mRNA expression of Rab1A in 176 of
tumor pancreatic tissue samples and 30 pancreatic normal tissue samples. Samples were produced at the pathology clinic, Uppsala University Hospital (Uppsala Sweden) [19]. 84 patients with high Rab1A were alive and 92 patients were dead. On the other hand, all alive patients have low Rab1A expression. Log-rank test was used to evaluate the significance of the correlation between Rab1A mRNA expression and patient survival. The analysis results were considered to be significant when the value of $p$ was less than 0.05.

2.3 Amplification of Rab 1A in pancreatic cancer

Many previous studies reported that Rab1A protein is overexpressed in different cancers [20-27]. Yet, its relevant role in pancreatic cancer is still obscure. Hence, the aim of our study is to examine and check the mRNA level of Rab1A in 176 pancreatic tumor and 248 non-cancerous pancreatic samples.

3. Results and Discussion

3.1 Potential role of the new Kaplan-Meier model in the early pancreatic cancer prognosis

In recent years, scientists tried to develop new methods to help the prediction of a cancer kind before it develops symptoms. However, an accurate prognosis of a tumor is considered a very difficult task for the medical practitioners. As a result, a lot of analysis models and technologies have been developed for the previously mentioned purpose. Kaplan Meier has become a very popular analysis tool among medical researchers. This analysis tool can predict the possible cancer outcomes by using very compound datasets. From using the basic features of the Kaplan-Meier classical model: gene symbol (Rab1A), follow up threshold (240 months), OS, cancer type (pancreatic cancer), we were able to predict the pancreatic tumor in later stages. Unexpectedly, analysis results revealed that Rab1A is amplified in pancreatic cancer patients. Moreover, the patient's survival rate was better with the low expression level of Rab1A in comparison to the poor survival of patients with a high expression level ($p=0.0024$) (Figure 1A). Besides our results are consistent with previously published papers and it’s overexpressed in many cancers [20-27]. Although classical features (as input) of the model can overall predict the possible outcomes of pancreatic cancer patients but the early prediction is remained not detected. So, we suggested in this study a new model with new restriction features (female gender, white race and follow up threshold 180 months) as input to improve the prognosis of human pancreatic cancer. Remarkably, the results of the new analysis model help to improve the detecting ability of the model in pancreatic cancer ($p<0.05$). The results are shown in Figure 1B. Our results demonstrate that high expression of Rab1A is linked to poor prognosis in white females’ stage 2 pancreatic tumors. The results are shown in Figure 1C. Together, our results show that the new model has better prognosis ability for the disease at early stages in comparison with the classical analysis model of Kaplan Meier. Furthermore, we further examined the impact of the new model on pancreatic cancer on earlier stage (stage 1). Our results are shown in Figure 1D. However, the results were not significant in terms of predicting the disease at earlier stages (stage 1) ($p>0.05$). Altogether, these observations reveal our improved analysis model could help in the early prediction of many further cancers.
A. Comparing the OS time of the two groups of pancreatic cancer patients, low and high Rab1A mRNA levels using the classical model, were p value was calculated using the log-rank test. B. The new model of Kaplan-Meier with new integrated features (180 months threshold, white race, and female gender) comparing the pancreatic cancer patient’s OS time. C. Improved Kaplan-Meier with four integrated features (180 months threshold, white as a race, female as gender, stage 2) comparing the OS time of pancreatic cancer patients. D. New model of Kaplan-Meier with four new integrated features (180 months threshold, white race, female gender, stage 1).

**Figure 1.** Potential role of the new Kaplan-Meier model in the early pancreatic cancer prognosis.
3.2 Validation the potential role of the Kaplan-Meier model in detection pancreatic cancer on the benchmark datasets from The Human Protein Atlas

Pancreatic cancer is one of the most aggressive cancer types. This reason makes finding the right treatment option for pancreatic cancer patients’ hard task for doctors. Therefore, we confirm the results of Kaplan model on the standard datasets deposited in The Human Protein Atlas [19]. Our Analysis for genomic datasets reveals that poorer survival is for pancreatic cancer patients with overexpression of Rab1A. While patients with low expression of Rab1A have better survival (p = 0.048) (Figure 2). Therefore, Rab1A level in pancreatic cancer patients can provide predictive value for pancreatic cancer patient’s outcomes. Our observations bear a resemblance to those results observed in our study and other studies [20-27]. The model results highly support that amplification of Rab1A is linked to a high risk of pancreatic cancer related death.

Figure 2. Validation the potential role of Kaplan-Meier model in prediction pancreatic cancer on benchmark datasets from The Human Protein Atlas.

Human Protein Atlas database used to validate our model of pancreatic cancer patients. Log rank test was used to calculate P value.

3.3 New Kaplan-Meier analysis model supports Rab1A role in oncogenesis

Several researches in the field of oncology believe Rab1A play a critical role in many tumors [20-27]. Here, the overexpression role Rab1A in pancreatic tumors was examined. The analysis results revealed that the level of mRNA in pancreatic tumors was higher than in noncancerous pancreatic samples. The results are depicted in Figure 3A and 3B. Generally, our results showed overexpression of Rab1A in pancreatic cancer similar to other studies [20-27]. Therefore, we believed that our results may elucidate several findings related to pancreatic cancer development.
Figure 3: Overexpression of Rab1A mRNA in pancreatic cancer. A. representation of Rab1A mRNA level in 248 pancreatic normal tissues. B. RNA seq data of Rab1A. Shown is mRNA level in 176 samples of pancreatic cancer tissues. The following link (https://www.proteinatlas.org/ENSG00000138069-RAB1A/pathology/pancreatic+cancer) was used for downloading data. FPKM is the fragment’s number of reads.

4. Conclusion
A new and improved model of Kaplan-Meier was presented in this work. An assessment for predicting pancreatic tumors at first and second stages in female white pancreatic tumor patients model was made for the model. Kaplan-Meier new model exhibits a high sensitivity for detecting the cancer in second stage in comparison to the classic one. Furthermore, the results were verified to a standard dataset from Human Protein Atlas for the pancreatic tumor prediction tasks [19]. Interestingly, our Kaplan Meier results were bearing a resemblance to standard datasets results. Overall, our results further emphasize the relevance of Rab1A in the development of other malignancies but its role in cancer development still not clear. So, more studies address Rab1A regulation in a pancreatic tumor are urgently required.

Acknowledgments
Many thanks go to Human Protein Atlas members for their valuable advice and assistance.

5. References
[1] Roth, Marc T et al. “Recent advances in the treatment of pancreatic cancer.” F1000Research vol. 9 F1000 Faculty Rev-131 (2020).
[2] McGuigan, Andrew et al. “Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes.” World journal of gastroenterology vol. 24,43: 4846-4861 (2018).
[3] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis. Computational and Structural Biotechnology Journal 13 8–17 (2015).
[4] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informat;2:59 (2006).
[5] Fortunato O, BoeriM, Verri C, Conte D,MensahM, Suatoni P, et al. Assessment of circulating microRNAs in plasma of lung cancer patients. Molecules; 19:3038–54 (2014).
[6] Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer. Curr Opin Pharmacol; 10:543–50 (2010).
[7] Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. Front Genet;4 (2013).
[8] Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev; 32:326–48 (2012).

[9] Silveira, Margarida, Jacinto C. Nascimento, Jorge S. Marques, André RS Marçal, Teresa Mendonça, Syogo Yamauchi, Junji Maeda, and Jorge Rozeira. "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images." IEEE Journal of Selected Topics in Signal Processing 3, no. 1: 35-45(2009).

[10] Cicchetti D. Neural networks and diagnosis in the clinical laboratory: state of the art. Clin Chem;38:9–10(1992).

[11] Cochran AJ. Prediction of outcome for patients with cutaneous melanoma. Pigment Cell Res;10:162–7 (1997).

[12] Exarchos KP, Goletsis Y, Fotiadis DI. Multiparametric decision support system for the prediction of oral cancer reoccurrence. IEEE Trans Inf Technol Biomed;16: 1127–34(2012).

[13] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med;23:89–109 (2001).

[14] Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. Engl Appl Artif Intell; 26:2194–205 (2013).

[15] Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. Bioinformatics;23:30–7(2007).

[16] Li Shen, Laurie R. Margolies, Joseph H. Rothstein, Eugene Fluder, Russell McBride and Weiva Sieh. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Scientific Reports 9:12495(2019).

[17] Kittler, Harold, H. Pehamberger, K. Wolff, and M. Binder. "Diagnostic accuracy of dermoscopy." The lancet oncology 3, no. 3: 159-165(2002).

[18] Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53, 457-481.

[19] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Šojžedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigiarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzez K, Forsberg M, Persson L, Johansson F, Zwaalenn M, von Heijne G, Nielsen J, Pontén F. Tissue-based map of the human proteome. Science 347(6220):1260419 (2015).

[20] Thomas JD, Zhang YJ, Wei YH, Cho JH, Morris LE, Wang HY, Zheng XF. Rab1A is an mTORC1 activator and a colorectal oncogene. Cancer Cell. ; 26:754–69(2014).

[21] Xu BH, Li XX, Yang Y, Zhang MY, Rao HL, Wang HY, Zheng XF. Aberrant amino acid signaling promotes growth and metastasis of hepatocellular carcinomas through Rab1A-dependent activation of mTORC1 by Rab1A. Oncotarget. ; 28(6):20813-28(2015).

[22] Bao ZS, Li MY, Wang JY, Zhang CB, Wang HJ, Yan W, Liu YW, Zhang W, Chen L, and Jiang T. Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling. CNS Neurosci Ther. ;20:112–18(2014).

[23] Abd Elmageed ZY, Yang Y, Thomas R, Ranjan M, Mondal D, Moroz K, Fang Z, Rezk BM, Moparty K, Sikk SC, Sartor O, and Abdel-Mageed AB. Neoplastic reprogramming of patient-derived adipose stem cells by prostate cancer cell-associated exosomes. Stem Cells. ;32:983–97(2014).

[24] Sun T, Wang X, He HH, Sweeney CJ, Liu SX, Brown M, Balk S, Lee GS, and Kantoff PW. MiR-221 promotes the development of androgen independence in prostate cancer cells via downregulation of HECTD2 and Rab1A. Oncogene. ;33:2790–800(2014).

[25] Shimada K, Uzawa K, Kato M, Endo Y, Shiiba M, Buka H, Yokoe H, Seki N, and Tanzawa H. Aberrant expression of RAB1A in human tongue cancer. Br J Cancer. ; 92:1915–21(2005).

[26] Zwyeya S. The role of overexpression of Rab1A in human pancreatic cancer and its association with poor prognosis. Indian journal of public health research and development; 10(11): 1854-1859 (2019).
[27] S Almansouri and S Zwyea Early prognosis of human renal cancer with Kaplan-Meier plotter data analysis model. Journal of physics: conference series 1530(1):012051.