Indonesian Journal of Medical Chemistry and Bioinformatics

Volume 1
Number 2 *Biomarker Metabolite Discovery for Pancreatic Cancer using Ma-chine Learning*

Article 4

3-20-2023

Biomarker Metabolite Discovery for Pancreatic Cancer using Machine Learning

Immanuelle Kezia
Universitas Indonesia, immanuelle.kezia@ui.ac.id

Linda Erlina
Universitas Indonesia, linda.erlina22@ui.ac.id

aryo tedjo
Universitas Indonesia, 1aryo.tedjo@gmail.com

Fadilah Fadilah
Universitas Indonesia, fadilah.msi@ui.ac.id

Biomarker Metabolite Discovery for Pancreatic Cancer using Machine Learning

Acknowledgements

We would like to say thank you for the helping of Aryo Tedjo as a conceptor, Fadilah Fadilah as a supervisor, and Linda Erlina as a proofreader. Also we want to say thank you a lot for Xiamen University, Department of Electronic Science that has published their raw data of metabolite on the metabolomic workbench, so we can use it for further analysis.

INDONESIAN JOURNAL OF MEDICAL CHEMISTRY AND BIOINFORMATICS
Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia

Article

Biomarker Metabolite Discovery for Pancreatic Cancer using Machine Learning

Immanuelle Kezia [1*], Linda Erlina [1,2], Aryo Tedjo [1,2], Fadilah Fadilah [1,2],

[1] Master's Programme in Biomedical Sciences, Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Indonesia
[2] Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Indonesia
* Correspondence: immanuelle.kezia@ui.ac.id

Abstract: Pancreatic cancer is one of the deadliest cancers in the world. This cancer is caused by multiple factors and mostly detected at late stadium. Biomarker is a marker that can identify some diseases very specific. For pancreatic cancer, biomarker has been recognized using blood sample known as liquid biopsy, breath, pancreatic secret, and tumor marker CA19-9. Those biomarkers are invasive, so we want to identify the disease using a very convenient method. Metabolite is product from cell metabolism. Metabolites can become a biomarker especially from difficult diseases. In this paper, we want to find biomarker from metabolite using machine learning and enrichment. Metabolites data was obtained from Metabolomic workbench, while the detection and identification is done using in silico. From 106 samples, control and cancer, we found 61 metabolites and analyze them. We got 8 metabolites that play important role in pancreatic cancer and found out 2 of them are the most impactful. From that we found that ethanol is one of the best candidate of biomarker that we provide for pancreatic detection cancer. However, the simulation need to be improved to find another biomarker that provide a better marker for prognosis.

Keywords: Metabolite; Pancreatic; Cancer; Machine Learning

Citation: Kezia, I.; Erlina, L.; Tedjo, A.; Fadilah, F. Biomarker Metabolite Discovery for Pancreatic Cancer using Machine Learning . *Ind. J. Med. Chem. Bio. IJMCB.* **2023**, 1, 2.

Received: Wed Jan 25, 2023
Accepted: Mon Mar 13, 2023
Published: Mon Mar 20, 2023

Publisher's Note: IJMCB stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The pancreas is an organ that produces enzymes that aid digestion as well as hormones that regulate blood sugar [1]. There are two types of growths in the pancreas: cancerous and non-cancerous [2]. Pancreatic cancer typically develops in cells found in the pancreatic duct, which transports digestive enzymes from the pancreas [2]. If there is a mutation and abnormal cell growth, the condition is known as pancreatic ductal adenocarcinoma (PDAC) [3].

Pancreatic cancer is the leading cause of death worldwide. This is due to the fact that patients with pancreatic cancer do not exhibit symptoms until the cancer has progressed to the stage of malignancy [4]. Men account for the majority of pancreatic cancer patients, and the disease's prevalence rises with age [5]. Pancreatic cancer is also inextricably linked to dysbiosis of the gut microbiota. *Pofphyromonas gingivalis* and *Granulicatella adiacen* populations are decreasing in pancreatic cancer patients. This population decline raises the likelihood of developing cancer [6], [7]. Furthermore, diabetes is the leading risk factor for pancreatic cancer. This is due to the pancreas producing the hormone insulin, which converts glucose into glucagon. When the pancreas grows abnormally, the production of the hormone insulin is disrupted, causing blood sugar levels to rise rapidly and lead to diabetes [3], [4].

Pancreatic cancer is classified into four types based on its clinical stage: type I, which occurs only in the pancreas and measures 2-4 cm, type II, which occurs in lymph nodes larger than 4 cm, and type III, which spreads throughout the lymphatic system. Type IV,

DOI: https://doi.org/10.7454/ijmcb.v1i2.1017. E-ISSN : 2963-3818

which has spread to other organs, affects the nervous and circulatory systems. Pancreatic cancer can generally only be identified when it is in type III or type IV, which is characterised by symptoms such as a yellow body colour change, drastic weight loss, abdominal pain, and easy fatigue [8], [9].

Biomarkers for pancreatic cancer detection have been developed, beginning with liquid biopsy, exhalation, and pancreatic secretion, to aid in diagnosis and more targeted therapy. The serum cancer antigen 19-9 (CA 19-9) is the most commonly used biomarker and has been approved by the FDA. The CA19-9 marker's low positive predictive value serves a purpose and has significance for monitoring treatment response and as a marker of recurrence [6]. Metabolites produced by cancer tissue can also be used to detect cancer and its recurrence [7], [10]. This metabolite is detected using spectrophotometric analysis and other similar technologies. Other biomarkers, such as volatile organic compounds found in respiration, are collected in a non-invasive manner. Sequencing is used to create biomarkers that can detect the presence of mutations in pancreatic secretions [11]. Therefore, this paper would like to predict biomarker from metabolite and validated it using machine learning.

2. Results

In this paper we used metabolites data from Metabolomic workbench using Project ID PR001339 and Study ID ST002113. Metabolite data was obtained from human blood and was characterized using NMR method. The search for biomarkers of these metabolites is intended to result in a diagnostic kit that will be used by humans in the future. As a result, the species Homo sapiens is used to avoid data bias.

NMR (Nuclear Magnetic Resonance) is one method for detecting metabolites. The primary goal of NMR is to detect changes in the atomic nucleus spin orientation [12]. NMR is extremely useful in determining the structure of a compound or metabolite. This is because NMR can detect interactions between protons. Chemical shift can occur because they are influenced by their surroundings. NMR is the most recent method for determining the structure of metabolites because it can distinguish the structure of isomers [2], [12].

Metabolomic workbench data were then prepared for opening in the orange analysis window. The sample must be in the column section, while the metabolites must be in the row section, according to the data criteria. Figure 1 depicts a data processing step performed with orange software.

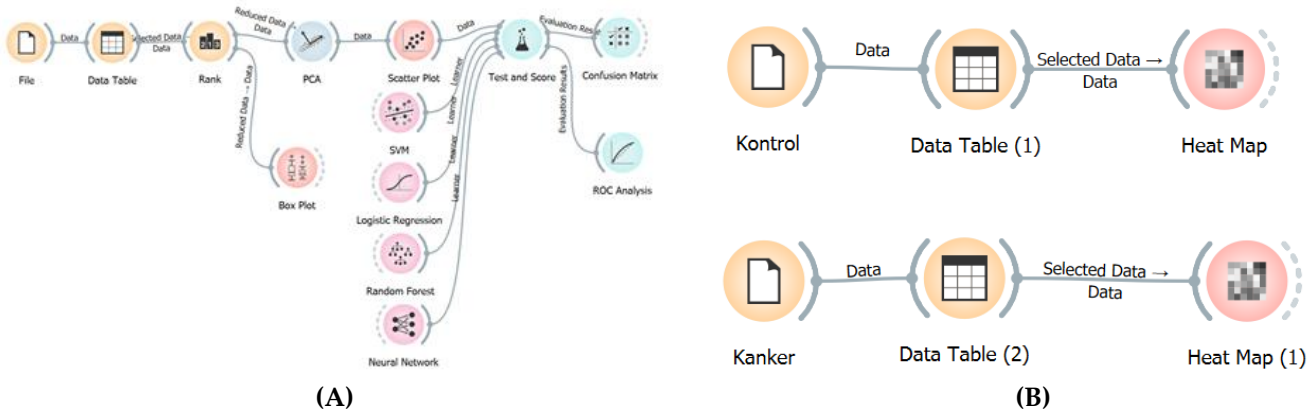

Figure 1. Data processing steps in orange (A) Machine learning processing and validation steps (B) Heatmap data processing steps.

PCA (Principle Component Analysis) is a technique for extracting patterns from large amounts of complex data. The entered dataset will be transformed into two datasets: one with individual weights and one with principal component weights [13]. Figure 2 depicts graph analysis components in the form

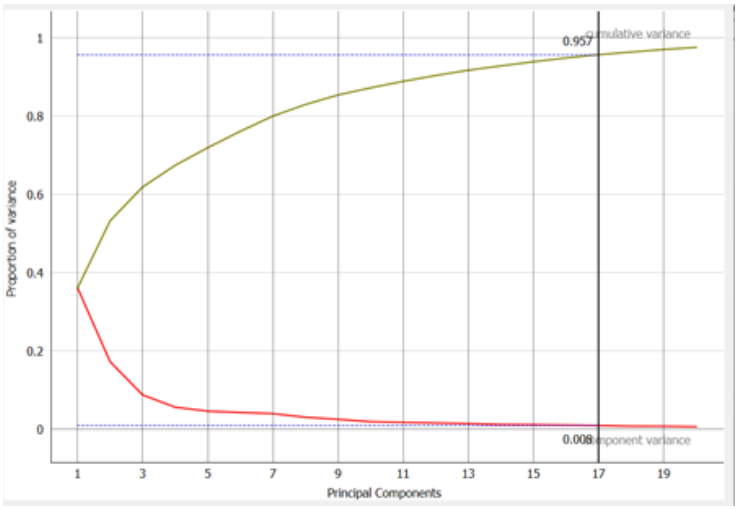

Figure 2. Srcee diagram of pancreatic cancer with controls

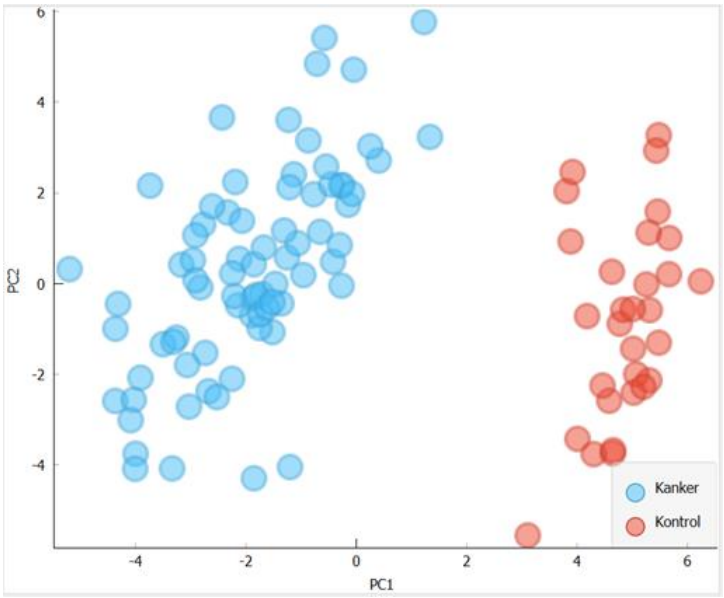

Figure 3. Scatter Plot of Pancreatic Cancer Metabolomics Dataset with Controls

In Figure 3, there is a clear distinction between PC1 and PC2. The dot graph also show the very significant separation among patients with pancreas cancer and control. It can be demonstrated that there is a difference in metabolite between pancreas cancer patients and control patients. Validation of these differences can be accomplished through the use of machine learning. Machine learning techniques used this year include SVM, random forest, neural networks, and logistic regression. The results of machine learning validation can be seen in Figure 4.

Figure 4 shows that the accuracy, specificity, and sensitivity are all greater than 100%. The AUC had an impact on this as well. AUC of 70% indicates that the data can be accurately, specifically, and sensitively analyzed and compared. The ROC graph supports this, with y axis representing sensitivity and x axis representing specificity. If the sensitivity rises while the specificity falls, the outcome will improve [21]. The ROC is set to 1 in Figures 4 (f), (g), (h), and I, indicating high sensitivity but low specificity.

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000	1.000
Random Forest	1.000	1.000	1.000	1.000	1.000
Neural Network	1.000	1.000	1.000	1.000	1.000
Logistic Regression	1.000	0.991	0.991	0.991	0.991

(A)

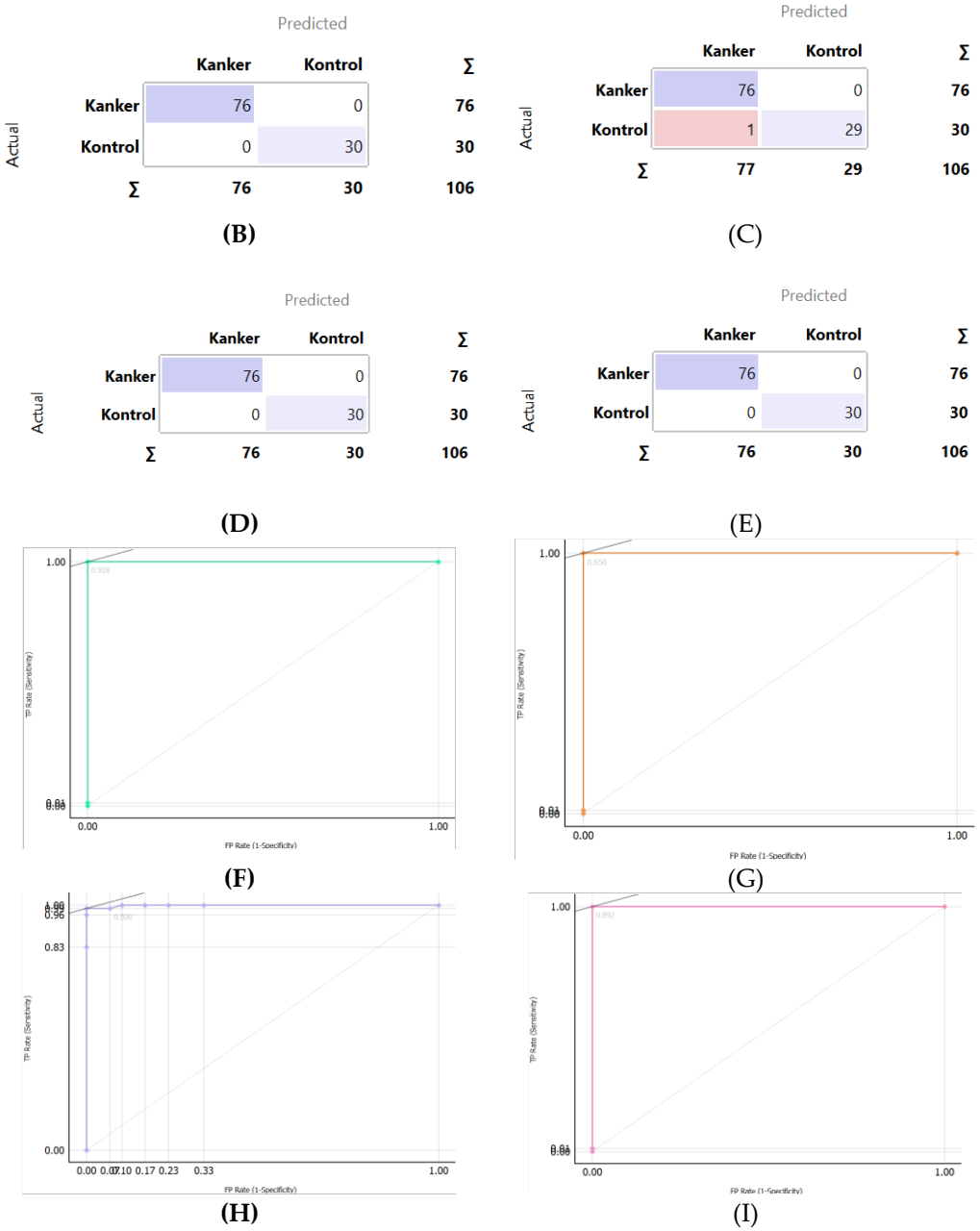

Figure 4. Validation of Machine Learning (A) Validation of Machine Learning (B) Confussion matrix SVM (C) Confussion matrix logistic regression (D) Confussion matrix random forest (E) Confussion matrix neural network (F) ROC curve SVM (G) ROC Curve Logistic Regression (H)ROC Curve Random Forest (I) ROC Curve Neural Network

According to the results of the statistical data analysis, metabolite in patients with pancreas cancer can be compared to metabolite control. From 106 metabolite data, 31 metabolite were detected in pancreas cancer. This metabolite can stimulate 81 genes associated with pancreas cancer. Metabolite and gen can be found in the table below.

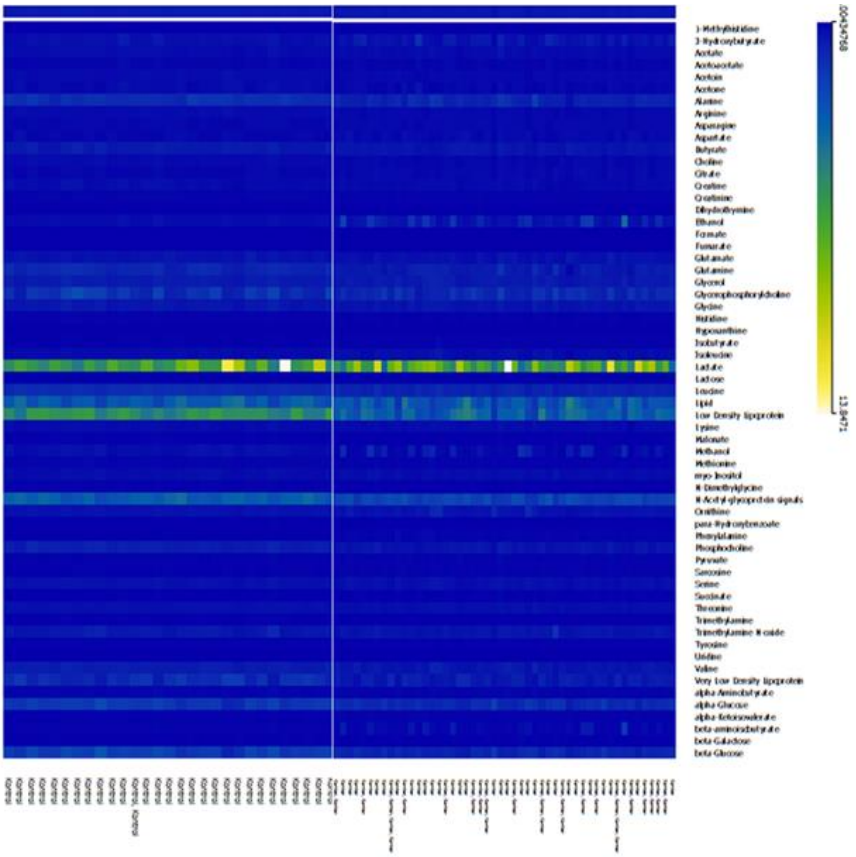

Figure 5. Heatmap of patients with normal pancreatic cancer.

3. Discussion

A scree diagram is a diagram that is used to determine the success of PCA in a dataset. The main components will be created based on the number of possible variations. PC1 will cover more variants than PC2 in the hierarchy. A scree diagram can assist in determining how much variation each PC (principal component) can capture. The y-axis represents the eigenvalues, while the x-axis represents the principal components. The eigenvalues will show how much variation there is. The ideal curve is steep and then sharply bends; the bend is the point of intersection, and the curve flattens. In Figure 2, the red line represents the variance covered per component, while the green line represents the total variance covered by the components. The red line in figure 2 is a curved scree. The curve will drop sharply at first, then bend. This is due to the fact that the first group is usually highly variable. Group 2 will have less variability because it is not excessive, whereas group 3 and subsequent variations will have less variability, resulting in a flatter curve [14]. The metabolomic pancreatic cancer dataset with controls can be distinguished from the scree diagram in Figure 2. The obvious bend in the principal component 2 reflected by the red line demonstrates this. Figure 3 shows a scatter plot that depicts the difference between metabolomic pancreatic cancer and controls.

Machine learning is used to determine the accuracy and specificity of data that has been collected [15]. SVM (support vector machine) is an algorithm used to compress data in N-dimensional space using hyperplanes [16]. Logistic Regression is a type of machine learning that falls under the category of supervised learning. Logistic Regression is frequently used to predict the likelihood of two variables in a given situation [17]. Random forest is a classification method that examines several points of interest in a dataset. In

contrast to opinion polls, the random forest algorithm was used to generate opinion polls. This will have a negative impact on the outcome of the prediction. Random forest has a distinct advantage in that it performs simulations more quickly than other algorithms and can make decisions with high accuracy [18]. A neural network is an artificial intelligence (AI) method that teaches computers to think like humans. The human brain, which synergizes to create memories, inspired neural network learning. Neural networks can help solve complex problems with high sensitivity [19]. A confusion matrix is a table that is used to determine a piece of data's accuracy, specificity, and sensitivity [20].

3-hydroxybutyrate, ethanol, lactate, methanol, and phosphocholine are among the metabolites that have increased in pancreatic cancer patients. Alanine, glycine, and trimethylamine-n-oxide are among the metabolites that have decreased in pancreatic cancer patients. While KRAS, TP53, and CDKN2A are all genes involved in pancreatic cancer. BRCA2 is a gene that contributes to DNA damage.

KRAS mutations are a hallmark of pancreatic cancer. Mutations in the KRAS gene are most common in the early stages of pancreatic cancer. Mutations in TP53 and CDKN2A occur after the progression of pancreatic cancer and have the function of increasing the cancer's invasiveness. KRAS is a gene that encodes a Ras family GTPase protein. KRAS has two paralogues, HRAS and NRAS, both of which are involved in the progression of cancer. Pancreatic cancer contains up to 85% of KRAS mutations. The mutation is a missense mutation that occurs at three residue hotspots: G12, G14, and Q61. These mutations impair the activity and binding efficiency of Ras protein to GTP during the active phase. KRAS signalling pathways include cell growth, cytokines, and hormone receptors [9], [22].

TP53 is a gene that produces the p53 protein, which acts as a tumour suppressor. Mutations in TP53 are extremely common, and the result is cancer progression. One of the causes of cancer progression in pancreatic cancer is deletion of the p16 locus on chromosome 9. Other types of mutations that occur are missense mutations and point mutations. These mutations can cause cells to exit the cell cycle and cause an increase in inflammation. This allows the cells to continue to proliferate and survive in hypoxic environments. Cells with TP53 mutations will also experience reprogramming of the metabolites produced in order to receive as much energy as possible in order to support unlimited cell proliferation. Mutations in the p53 gene can also cause metastasis and inhibit autophagy [23]–[25].

The CDKN2A gene is yet another tumour suppressor gene. The CDKN2A gene is found on chromosome 9p21. Pancreatic cancer can be caused by both germline and somatic mutations in the CDKN2A gene. The CDKN2A gene can encode a tumour suppressor with three different exons. Its primary function is to prevent progression at cell cycle checkpoints during the G1 phase. Furthermore, CDKN2A inhibits the phosphorylation of retinoblastoma protein, which inhibits the E2F transcription factor. When CDKN2A is mutated, cancer progresses by inducing excessive cell proliferation [26].

The tumour suppressor gene BRCA2 is found on the autosomal chromosome. Mutations in the BRCA gene cause cells to lose control of their growth and differentiation, allowing tumorigenesis to begin. BRCA is also a protein that aids in the repair of DNA damage, which is typically caused by dsDNA damage. BRCA mutations in pancreatic cancer can be caused by lifestyle factors such as smoking and drinking alcohol. This is due to the fact that smoking and alcohol are carcinogens that can cause unrestricted cell proliferation [27], [28].

Ethanol play important role in many mutational cases. Ethanol is obtained by consumable of alcohol drink and glycolysis products. High-intensity ethanol exposure can promote tumour aggressiveness by activating invasive phenotypes and allowing tumours to migrate. Ethanol contributes to tumorigenesis by suppressing the immune system and increasing ROS levels. Alcohol can activate the Ras/MEK/MAPK signalling pathway. These signalling pathways have the potential to influence cell growth. This is due to the fact that metabolites produced by ethanol-exposed cells cause DNA mutations, altering

proteins involved in cell proliferation. Ethanol derived from alcohol consumption can alter KRAS transcription and translation processes. Continuous alcohol exposure can cause genomic instability and cytoskeleton disruption, allowing mutated cells to migrate elsewhere. Ethanol is thought to play a role in p53 mutase, resulting in unrestricted cell growth. Ethanol and BRCA2 also play a role in DNA repair, with ethanol induction causing BRCA2A to lose its ability to repair DNA. The resulting genomic instability causes excessive cell proliferation, resulting in a hypoxic environment in the tumour. The energy obtained during the process of continuous cell division comes from the breakdown of glycolytic sugars, which produces ethanol metabolites. This is why cancer patients have high ethanol levels [9], [29].

Tumors can live in hypoxic environments, which is one of their characteristics. A hypoxic environment is one that lacks oxygen and is acidic on the outside. Tumors have the ability to maintain an acidic internal environment. Hypoxia induced factor (HIF) is a protein that promotes cancer cell growth in hypoxic environments. HIF1-alpha has the ability to upregulate proteins required for cancer cell growth in pancreatic cancer. Increased HIF1-alpha expression leads to increased GLUT-1 expression, which is involved in glycolysis. Lactate is produced as a byproduct of the glycolysis process. Glycolysis occurs in normal cells, but ATP energy is produced via the elector transfer stage. All energy in cancer cells is produced through the glycolysis process. Furthermore, cancer cells require a lot of energy to proliferate and even invade. This causes glycolysis to be activated, resulting in the accumulation of lactate byproducts [30]–[32].

In cases of hypoxia, other energy sources, such as ketones, are used in addition to glucose. Ketone bodies are compounds that are produced as a result of fat metabolism. Acetone, acetoacetic acid, and beta-hydroxybutyrate are examples of these compounds. Ketone body metabolism can be activated in pancreatic ductal adenocarcinoma (PDA) cells. One of the metabolites produced by metabolism that can promote cancer cell growth is beta-hydroxybutyrate. PDA cells can activate enzymes involved in ketogenesis and use a variety of nutrient and carbon sources. Excess beta-hydroxybutyrate can promote uncontrolled cell proliferation and even metastasis [33].

The formation of new cells as a result of cell division undoubtedly necessitates a composition such as the plasma membrane. The phospholipid bilayer, which contains phospholipids that bind to choline, is one of the plasma membrane's constituents. Phospholipids are also referred to as phosphocholine. In cases of pancreatic cancer caused by mutations in several genes such as KRAS, CDKN2A, and TP53, the cells proliferate excessively. This raises phosphocholine levels in the body, which helps to form the plasma membrane. Phosphocholine levels that are high indicate active cell division [7], [34].

Alanine is an amino acid that is used in the production of proteins. Extracellular alanine is a critical component for T cell re-stimulation and activation. Alanine binds to the SNAT1 receptor and stimulates the production of proteins necessary for cell growth, proliferation, and cytokine production. This is due to the fact that when T cells are activated, effector cells must grow and proliferate quickly. To provide this energy, the activation of glucose and amino acid transporters, including the alanine transporter, is increased. Cancer has the ability to fool the immune system, which is one of its characteristics. Deception is accomplished by lowering alanine metabolites, preventing T cell reactivation. As a result, T cells do not recognise tumour cells, while tumor cells are able to proliferate and invade other tissues [2], [35].

Glycine is a non-essential amino acid that can be synthesised in the body. Glycine has the ability to conjugate with bile. The gut microbiota produce ursodeoxycholic acid (UCDA), a secondary bile capable of conjugating with glycine. The conjugation has the ability to increase hydrophilicity and thus reduce cytotoxicity. UDCA has antioxidant, anti-inflammatory, and cytoprotective properties. These properties can inhibit cell proliferation by preventing it from entering the G1 phase. UDCA can also prevent Ras and Cox-2 mutations, which are key players in pancreatic cancer. This causes glycine to be produced in large quantities in normal cells while decreasing glycine production in cancer cells [34].

The gut microbiota produces trimethylamine N-oxide (TMAO), a choline metabolite. The enzyme flavin-containing monooxygenase-3 aids in the production of TMAO in the liver (FMO3). TMAO and ROS levels appear to be elevated. Through the PERK signalling pathway, TMAO can cause apoptosis. Misfolding of the proteins that make up the ER causes stress on the ER in some conditions. This will activate several proteins, including ATF6. ATF6 has the ability to translocate to the Golgi apparatus and split into 50 kDa protein fragments. After that, the fragment enters the nucleus and binds to nuclear factor Y. This complex then binds to the ER stress response (ERSE) and causes the chaperone and XBP1 to be expressed. This will cause the induction of making H2O2 and also O2 due to the interaction of the UPR with mitochondria. In this case, TMAO was able to induce apoptosis and oxidative stress via the IRE1-alpha/XBP-1 signaling pathway [36], [37].

4. Materials and Methods

4.1 Obtaining Metabolite Data from Metabolomic Workbench

The Metabolomic Workbench could be found at https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Project&ProjectID=PR001339, accessed on October 10th, 2022. The keyword pancreatic cancer would be entered in the query section. Following that, the appropriate data search was performed, where it could be seen in the "view" section whether there was metabolite data or only meta data. To reduce the occurrence of bias, it was also necessary to sort the species that were chosen from humans rather than test animals in the metadata section. The time of data collection was also an important parameter to consider because the more recent the year of publication, the more recent the technology used, resulting in a renewable metabolite profile.

4.2 Biomarker Data Analysis using Orange Software

The data obtained from the metabolomic workbench is in .txt. The data must be converted to.xls format before it could be transferred and analyzed in orange software version 3.31.1. Sample data should be organized in columns, while metabolic data should be organized in rows. The data was then imported into the orange software, and the data was normalized before the analysis begins. The normalized data was then ranked in order, followed by PCA analysis. PCA visualization was accomplished through the use of scatter plots and heatmaps.

4.3 Machine Learning Validation using Orange Software

Machine learning was used to validate orange data that had been analyzed. SVM, logistic regression, and neural algorithms are examples of machine learning. The AUC curve and the confusion matrix were visible parameters. Data validation would be obtained from the machine learning algorithm by examining the sensitivity and specificity.

4.4 Correlation of Metabolites and Genes using CTD

The CTD web server (http://ctdbase.org/tools/batchQuery.go, accessed on October 25th, 2022) was then used to enrich the metabolite data. A list of metabolites would be entered in the query section, and a search would be performed. Because the data from the CTD was in.txt format, it must be converted into an.xls file for analysis.

4.5 Correlation of Metabolites and Phenotype using CTD

Metabolite data could also be analysed using the CTD web server (http://ctdbase.org/tools/batchQuery.go, accessed on October 25th, 2022) and Metaboanalyst (https://www.metaboanalyst.ca/MetaboAnalyst/, accessed on October 30th, 2022). The disease category was chosen, and Metaboanalyst enrichment was chosen. On CTD, the data will be in the form of.txt, whereas on Metaboanalyst, the data would be in the form of.png.

5. Conclusions

Based on the result, we know that metabolite biomarker for pancreatic cancer are 3-hydroxybutyrate, ethanol, lactate, methanol, phosphocoline, alanine, glycine, and trimethylamine N-oxide. This metabolite chances can be different from each person due to life style and medication. Because of those limitation, in vitro and in vivo studies should be conduct to prove and make sure that those metabolites are changes in pancreatic cancer patients.

Acknowledgement: We would like to say thank you for the helping of Aryo Tedjo as a conceptor, Fadilah Fadilah as a supervisor, and Linda Erlina as a proofreader. Also we want to say thank you a lot for Xiamen University, Department of Electronic Science that has published their raw data of metabolite on the metabolomic workbench, so we can use it for further analysis.

References

1. [A. McGuigan, P. Kelly, R. C. Turkington, C. Jones, H. G. Coleman, and R. S. McCain, "Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes," *World J. Gastroenterol.*, vol. 24, no. 43, pp. 4846–4861, 2018, doi: 10.3748/wjg.v24.i43.4846.
2. Y. J. McConnell *et al.*, "Distinguishing benign from malignant pancreatic and periampullary lesions using combined use of1h-nmr spectroscopy and gas chromatography–mass spectrometry," *Metabolites*, vol. 7, no. 1, pp. 1–15, 2017, doi: 10.3390/metabo7010003.
3. P. Rawla, T. Sunkara, and V. Gaduputi, "Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors," *World J. Oncol.*, vol. 10, no. 1, pp. 10–27, 2019, doi: 10.14740/wjon1166.
4. J. M. Winter, A. Maitra, and C. J. Yeo, "Genetics and pathology of pancreatic cancer," *Hpb*, vol. 8, no. 5, pp. 324–336, 2006, doi: 10.1080/13651820600804203.
5. "Pancreatic Cancer—Patient Version - NCI." 2018. Retrieved on 11th May 2022.
6. R. Rikarni, "Pancreatic Cancer: Pathogenesis, Diagnosis, and Laboratory Tests," *Indones. J. Clin. Pathol. Med. Lab.*, vol. 27, no. 3, pp. 333–340, 2021, doi: 10.24293/ijcpml.v27i3.1891.
7. J. Mayerle *et al.*, "Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis," *Gut*, vol. 67, no. 1, pp. 128–137, 2018, doi: 10.1136/gutjnl-2016-312432.
8. R. A. C. M. Boonen, M. P. G. Vreeswijk, and H. van Attikum, "Functional Characterization of PALB2 Variants of Uncertain Significance: Toward Cancer Risk and Therapy Response Prediction," *Front. Mol. Biosci.*, vol. 7, no. September, 2020, doi: 10.3389/fmolb.2020.00169.
9. J. G. T. 3 Emalie J. Clement 1, Henry C.-H. Law 1 , Fangfang Qiao 1 , Dragana Noe 2 and and N. T. Woods, "Combined Alcohol Exposure and KRAS Mutation in Human Pancreatic Ductal Epithelial Cells Induces Proliferation and Alters Subtype Signatures Determined by Multi-Omics Analysis," *Cancers (Basel)*, vol. 14, no. 8, p. 1968, 2022, doi: 10.3390/cancers14081968.
10. Y. Wang *et al.*, "Prognostic Biomarkers for Pancreatic Ductal Adenocarcinoma: An Umbrella Review," *Front. Oncol.*, vol. 10, no. September, pp. 1–12, 2020, doi: 10.3389/fonc.2020.01466.
11. Y. Cao *et al.*, "Potential Metabolite Biomarkers for Early Detection of Stage-I Pancreatic Ductal Adenocarcinoma," *Front. Oncol.*, vol. 11, no. January, pp. 1–10, 2022, doi: 10.3389/fonc.2021.744667.
12. C. Dalvit, E. Ardini, M. Flocco, G. P. Fogliatto, N. Mongelli, and M. Veronesit, "A General NMR Method for Rapid, Efficient, and Reliable Biochemical Screening," *J. Am. Chem. Soc.*, vol. 125, no. 47, pp. 14620–14625, 2003, doi: 10.1021/ja038128e.
13. I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
14. R. D. Ledesma, P. Valero-Mora, and G. Macbeth, "The Scree Test and the Number of Factors: a Dynamic Graphics Approach," *Span. J. Psychol.*, vol. 18, no. June, p. E11, 2015, doi: 10.1017/sjp.2015.13.
15. P. Larrañaga *et al.*, "Machine learning in bioinformatics," *Brief. Bioinform.*, vol. 7, no. 1, pp. 86–112, 2006, doi: 10.1093/bib/bbk007.
16. H. Ahmadi, P. Pichappan, and E. Ariwa, *Communications in Computer and Information Science: Preface*, vol. 241 CCIS. 2011.
17. C. Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002, doi: 10.1080/00220670209598786.
18. A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.
19. F. Dhimas Syahfitra, R. Syahputra, and K. Trinanda Putra, "Implementation of Backpropagation Artificial Neural Network as a Forecasting System of Power Transformer Peak Load at Bumiayu Substation," *J. Electr. Technol. UMY*, vol. 1, no. 3, pp. 118–125, 2017, doi: 10.18196/jet.1316.
20. D. Visa Sofia, "Confusion Matrix-based Feature Selection Sofia Visa," *ConfusionMatrix-based Featur. Sel. Sofia*, vol. 710, no. January, p. 8, 2011.
21. K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Casp. J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.
22. J. Luo, "KRAS mutation in pancreatic cancer," *Semin. Oncol.*, vol. 48, no. 1, pp. 10–18, 2021, doi: 10.1053/j.seminoncol.2021.02.003.
23. I. A. Voutsadakis, "Mutations of p53 associated with pancreatic cancer and therapeutic implications," *Ann. Hepato-Biliary-Pancreatic Surg.*, vol. 25, no. 3, pp. 315–327, 2021, doi: 10.14701/ahbps.2021.25.3.315.

24. Y. Tabach *et al.*, "Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer," *PLoS One*, vol. 6, no. 1, 2011, doi: 10.1371/journal.pone.0014632.
25. K. Ando *et al.*, "Discrimination of p53 immunohistochemistry-positive tumors by its staining pattern in gastric cancer," *Cancer Med.*, vol. 4, no. 1, pp. 75–83, 2015, doi: 10.1002/cam4.346.
26. R. R. McWilliams *et al.*, "Prevalence of CDKN2A mutations in pancreatic cancer patients: Implications for genetic counseling," *Eur. J. Hum. Genet.*, vol. 19, no. 4, pp. 472–478, 2011, doi: 10.1038/ejhg.2010.198.
27. J. B. Greer and D. C. Whitcomb, "Role of BRCA1 and BRCA2 mutations in pancreatic cancer," *Gut*, vol. 56, no. 5, pp. 601–605, 2007, doi: 10.1136/gut.2006.101220.
28. Human Prostate Gene Database, "BRCA1 BRCA1 DNA repair associated [Homo sapiens (human)] - Gene - NCBI." 2002, [Online]. Available: https://www.ncbi.nlm.nih.gov/gene/672, accessed on November 4th, 2022.
29. F. Harish Lavu, MD, FACS, Harry B Lengel, BA, Naomi M Sell, BA, Joseph A Baiocco, BS, Eugene P Kennedy, MD, FACS, Theresa P Yeo, PhD, Sherry A Burrell, PhD*, Jordan M Winter, MD, FACS, Sarah Hegarty, MPhil^, Benjamin E. Leiby, PhD^, and Charles J Yeo, MD, "A Prospective randomized double blind placebo controlled trial on the efficacy of ethanol cellac Plexsus neurolysis in patients with operable panceratic and periampullary adenocarcinoma," *J Am Coll Surg*, vol. 220, no. 4, pp. 497–508, 2015, doi: 10.1016/j.jamcollsurg.2014.12.013.
30. B. Muz, P. de la Puente, F. Azab, and A. K. Azab, "The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy," *Hypoxia*, p. 83, 2015, doi: 10.2147/hp.s93413.
31. Y. Xiao *et al.*, "Prognostic relevance of lactate dehydrogenase in advanced pancreatic ductal adenocarcinoma patients," *BMC Cancer*, vol. 17, no. 1, pp. 1–7, 2017, doi: 10.1186/s12885-016-3012-8.
32. M. E. Cameron, A. Yakovenko, and J. G. Trevino, "Glucose and Lactate Transport in Pancreatic Cancer: Glycolytic Metabolism Revisited," *J. Oncol.*, vol. 2018, 2018, doi: 10.1155/2018/6214838.
33. S. K. Shukla *et al.*, "Erratum to: Metabolic reprogramming induced by ketone bodies diminishes pancreatic cancer cachexia," *Cancer Metab.*, vol. 2, no. 1, pp. 1–19, 2014, doi: 10.1186/2049-3002-2-22.
34. H. N. Luu *et al.*, "The Association between Serum Serine and Glycine and Related-Metabolites with Pancreatic Cancer in a Prospective Cohort Study," *Cancers (Basel).*, vol. 14, no. 9, 2022, doi: 10.3390/cancers14092199.
35. N. Ron-Harel *et al.*, "T Cell Activation Depends on Extracellular Alanine," *Cell Rep.*, vol. 28, no. 12, pp. 3011-3021.e4, 2019, doi: 10.1016/j.celrep.2019.08.034.
36. G. Yang and X. Zhang, "TMAO promotes apoptosis and oxidative stress of pancreatic acinar cells by mediating IRE1α-XBP-1 pathway," *Saudi J. Gastroenterol.*, vol. 27, no. 6, pp. 361–369, 2021, doi: 10.4103/sjg.sjg_12_21.
37. Z. Y. Liu *et al.*, "Trimethylamine N-oxide, a gut microbiota-dependent metabolite of choline, is positively associated with the risk of primary liver cancer: A case-control study," *Nutr. Metab.*, vol. 15, no. 1, pp. 1–9, 2018, doi: 10.1186/s12986-018-0319-2.