

URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION USING DATA SCIENCE TECHNIQUE

Submitted in partial fulfillment of the requirements for the award of
MASTER OF SCIENCE IN BIOINFORMATICS AND DATA SCIENCE

By

DHEENA VR (42739009)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A++” by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL - 2024

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A++" grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **DHEENA VR (42739009)**, who carried out the project entitled "**URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION USING DATA SCIENCE TECHNIQUE**" under our supervision from September 2023 to April 2024.



Dr. H. JEMMY CHRISTY
Internal Guide



Dr. H. JEMMY CHRISTY
External Guide



Dr. H. JEMMY CHRISTY
Head of the Department

Submitted for Viva Voce Examination held on 20/04/2024



Internal Examiner



External Examiner

DECLARATION

I, **DHEENA VR (42739009)**, hereby declare that the Project Report entitled **URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION USING DATA SCIENCE TECHNIQUE** done by me under the guidance of **Dr. H. JEMMY CHRISTY** at **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY, CHENNAI**, and is submitted in partial fulfillment of the requirements for the award of Master of Science in Bioinformatics and Data Science.

DATE: 20/04/2024

PLACE: CHENNAI



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENTS

I am pleased to acknowledge my sincere thanks to the **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project for completing it successfully. I am grateful to them.

I extend my sincere gratitude to **Dr. REKHA CHAKRAVARTHY, Dean of the School of Science and Humanities**, and I am deeply grateful for her mentorship and leadership, which have been a constant source of inspiration. Her dedication to fostering a conducive academic environment has been instrumental in the successful completion of this thesis.

I convey my thanks to **DR. H. JEMMY CHRISTY, Head of the Department, Department of Bioinformatics**, for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **DR. H. JEMMY CHRISTY** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project.

ABSTRACT

The quest for urinary biomarkers in pancreatic cancer detection represents a pivotal pursuit for enhancing early diagnosis and therapeutic interventions. Employing machine learning algorithms and statistical methodologies, researchers endeavour to pinpoint distinctive biomarkers in urine samples capable of reliably signalling the presence or progression of pancreatic cancer. Initial phases entail meticulous data collection, gathering urine specimens from pancreatic cancer patients and healthy cohorts for comparative analysis, often utilizing techniques like mass spectrometry or immunoassays to identify potential biomarkers. Subsequently, machine learning algorithms are trained using this data to construct predictive models discerning pancreatic cancer presence based on biomarker presence. Rigorous model evaluation ensues, involving comprehensive testing against independent datasets to gauge accuracy, sensitivity, specificity, and other performance metrics, alongside validation studies ensuring biomarker reliability across diverse populations. From data preprocessing to model development and evaluation, a systematic analytical framework is employed to determine the most accurate algorithm for practical application in pancreatic cancer diagnosis.

KEYWORDS: Pancreatic cancer, Biomarkers, Early Diagnosis, Disease progression, Data mining, Machine learning technique, Prediction

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	TABLE OF CONTENTS	ii
	LIST OF FIGURES	iv
	LIST OF ABBREVIATION	v
1	INTRODUCTION	
1.1	URINARY BIOMARKERS	1
1.2	PANCREATIC CANCER	1
1.3	BACKGROUND	2
2	REVIEW OF LITERATURE	3
3	AIM AND OBJECTIVES	
3.1	AIM	6
3.2	OBJECTIVES	6
3.3	SCOPE	6
4	MATERIALS AND METHODOLOGY	
4.1	WORKFLOW	7
4.2	DATA COLLECTION	7
4.2.1	KAGGLE DATASET	8
4.3	IMPORT LIBRARIES	8
4.4	PREPROCESSING THE URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING LABEL ENCODER	9
4.5	VISUALIZING THE URINARY BIOMARKER FOR PANCREATIC CANCER USING MATPLOTLIB	9
4.6	MODEL BUILT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER USING CATBOOST CLASSIFIER	10
4.6	MODEL BUILT FOR URINARY BIOMARKER FOR PANCREATIC CANCER USING RANDOM FOREST	11

	MODEL BUILT FOR URINARY BIOMARKER FOR PANCREATIC CANCER USING NAÏVE BAYES	
4.7	CLASSIFIER	13
4.8	DEPLOYMENT- DJANGO	14
5	RESULT AND DISCUSSION	
5.1	RESULT	17
	PREPROCESSING THE URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING LABEL	
5.1.1	ENCODER	17
	VISUALIZING THE URINARY BIOMARKER FOR	
5.1.2	PANCREATIC CANCER USING MATPLOTLIB	19
	MODEL BUILT FOR URINARY BIOMARKERS FOR	
5.1.3	PANCREATIC CANCER USING CATBOOST CLASSIFIER	23
	MODEL BUILT FOR URINARY BIOMARKER FOR	
5.1.4	PANCREATIC CANCER USING RANDOM FOREST	25
	MODEL BUILT FOR URINARY BIOMARKER FOR PANCREATIC CANCER USING NAÏVE BAYES	
5.1.5	CLASSIFIER	28
5.1.6	DEPLOYMENT	29
5.2	DISCUSSION	31
6	SUMMARY AND CONCLUSION	32
	REFERENCES	33

LIST OF FIGURES

FIGURE		PAGE
NO	TITLE	NO
4.1	WORKFLOW	7
5.1	PREPROCESSING THE URINARY BIOMARKERS FOR THE PANCREATIC CANCER DATASET	18
5.2	PIE PLOT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER	19
5.3	COMPARING DIAGNOSIS WITH SEX IN URINARY BIOMARKERS FOR PANCREATIC CANCER	20
5.4	COMPARING DIAGNOSIS WITH AGE IN URINARY BIOMARKERS FOR PANCREATIC CANCER	21
5.5	COMPARING SAMPLE ORIGIN VS DIAGNOSIS IN URINARY BIOMARKERS FOR PANCREATIC CANCER	21
5.6	COMPARING CRETANINE VS DIAGNOSIS IN URINARY BIOMARKERS FOR PANCREATIC CANCER	22
5.7	PAIR PLOT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER	23
5.8	ACCURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING CATBOOST CLASSIFIER	25
5.9	ACCURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING RANDOM FOREST	27
5.10	ACCURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING NAIVE BAYES CLASSIFIER	29
5.11	URINARYBIOMARKERS FOR PANCREATIC CANCER PREDICTION LOGIN PAGE	29
5.12	URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION WEBPAGE(A)	30
5.13	URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION WEBPAGE(B)	30

LIST OF ABBREVIATIONS

MRI	MAGNETIC RESONANCE IMAGING
CT	COMPUTED TOMOGRAPHY
PDAC	PANCREATIC DUCTAL ADENOCARCINOMA
SVM	SUPPORT VECTOR MACHINE
SVC	SUPPORT VECTOR CLASSIFIER
GUI	GRAPHICAL USER INTERFACE
VSCODE	VISUAL STUDIO CODE
HTML	HYPERTEXT MARKUP LANGUAGE
XML	EXTENSIBLE MARKUP LANGUAGE
SQL	STRUCTURED QUERY LANGUAGE
MVC	MODEL VIEW CONTROLLER

CHAPTER 1

INTRODUCTION

1.1 URINARY BIOMARKERS

The use of urinary biomarkers to diagnose disease is a long-standing practice. Ancient clinicians detected glucose in the urine by tasting it or observing whether it attracted ants. The presence of albumin in the urine has been measured as an indicator of renal disease for centuries and in early times could be detected via the so-called “foam test” to determine whether albumin was present in the urine in large amounts. Even today clinicians frequently shake a urine sample to determine whether it develops a froth, *prima facie* evidence for a high level of protein, which often is indicative of glomerular disease. In that tradition, studies to identify biomarkers of disease in the urine have been an underlying component of investigative medicine throughout the 20th century and the early 21st century. These studies have been based on knowledge of the pathophysiology of disease to identify putative biomarkers that could be tested in clinical trials (Yoon *et al.* 2024)

1.2 PANCREATIC CANCER

Pancreatic cancer is one of the most lethal malignant tumors, characterized by delayed diagnosis, difficult treatment, and high mortality. The overall five-year survival rate among patients is less than 9%. In the process of diagnosis and treatment, an accurate segmentation of pancreatic cancer plays an important role. Attiyeh *et al.* performed 3D modeling based on accurately segmented pancreatic cancer, calculated the tumor size, extracted texture features, and built survival prediction models for patients. Magnetic resonance-guided radiation therapy has potential advantages in treating locally advanced pancreatic cancer. The critical step in this method is contouring, whereby a target volume is generated at the beginning of each treatment stage. (Siahaan (2023)

1.3 BACKGROUND

Pancreatic cancer is a murderous nasty excrescence with one of the worst vaticinations. Accurate segmentation of pancreatic cancer is vital in clinical opinion and treatment. Due to the unclear boundary and small size of cancers, it's challenging to both manually annotate and automatically member cancers. Considering 3D information application and small sample sizes, we propose a model-driven deep literacy system for pancreatic cancer segmentation grounded on helical metamorphosis. This study developed a combined helical-metamorphosis and model-driven deep literacy system for pancreatic cancer segmentation. A new spiral transformation algorithm was developed to enable a 2D model to achieve 3D image segmentation with affordable computational coffers. It also handed colorful 2D transformed images to realize an effective data addition for easing the problem of inadequate samples. either, a metamorphosis-weight-corrected module was designed to combine with the 2D model in a unified frame for the 2D segmentation and 3D reconditioning. The expansive trials conducted on internal multi-parametric MRI datasets and an external CT dataset demonstrated that the proposed system exhibits promising segmentation performance with strong robustness and generality while reaching a good balance between the performance and the computational coffers. (Cranfield (2023))

CHAPTER 2

REVIEW OF LITERATURE

Julius M. Kernbach, and Victor E. Staartjes *et al* (2020) Predicted Prognosis of Pancreatic Cancer Patients by Machine Learning: We recently read the publication by Yokoyama and colleagues, in which the scientists present a prognostic model that forecasts overall survival in patients with pancreatic cancer at a given 5-year interval by integrating the DNA methylation state of three mucin genes. They collected samples from 191 patients and compared support vector machines (SVM) with several kernels of different levels of complexity against a neural network. Models were trained using a 50/50 split, leave-one-out cross-validation, or k-fold cross-validation (with varying values for k) to evaluate generalizability. Nevertheless, neither the testing set nor the training set provided information on typical discriminative performance metrics like sensitivity, specificity, AUC, or F1-Score; The model was not calibrated; resampling was used inconsistently; and model selection and evaluation were not done independently.

Ahmad R. Cheema, MD^{a,b}, Eileen M. O'Reilly, MD^{c,d} *et al* (2016) Management of Metastatic Pancreatic Adenocarcinoma: It is anticipated that 53,000 new cases of PDAC diagnoses will occur in the US in 2016. This corresponds to about 3% of all cases of cancer. In the United States, pancreatic cancer ranks fourth for cancer-related deaths in both men and women, with a high mortality-to-incidence ratio and a low incidence. It is expected to move up to the second place for fatalities from cancer by 2030. PDAC is the most common histologic subtype, accounting for about 85% of cases. A tiny proportion of patients benefit greatly from surgical resection; nevertheless, at the time of diagnosis, more than 50% of patients had distant metastases, and 70% to 80% of patients had an incurable condition.

Dimitrios Giannis, Dimitrios Moris, and Andrew S. Barbas *et al* (2021) Diagnostic, Predictive and Prognostic Molecular Biomarkers in Pancreatic Cancer: An Overview for Clinicians: The most prevalent pancreatic cancer, known as pancreatic ductal adenocarcinoma (PDAC), has a dismal prognosis and an aggressive tumor behavior. Treatment-resistant tumors and advanced disease stages are common in PDAC patients. The search for novel biomarkers is necessary because there are currently no noninvasive tests available for PDAC diagnosis and survival prediction. Individuals with PDAC and those at high risk must be identified early. Furthermore, it is essential to identify the chemicals linked to tumor biology, aggressiveness, and metastatic potential to forecast survival and offer patients individualized treatment plans.

Santosh Reddy P *et al* (2020) A Survey On Prediction Of Survival Time On Pancreatic Cancer Using Machine Learning Paradigms Towards Big Data: AI and quantifiable multi-variate relapses are two examples of paradigms developed through several prior methods that can be used to build resistance to an imminent sickness. In any event, not every situation may benefit from such a strategy's most predictive representations. to serve as a representation of automated meta-training techniques that specify how to forecast the optimal system performance for every patient. The highly selective process is employed to sustain the patient's resistance. We assessed the suggested methods using a database of meticulous pancreatic disease resection review records.

E. P. DiMagno *et al* (1999) Pancreatic cancer: Clinical presentation, pitfalls and early clues: Since pancreatic cancer is usually diagnosed based on symptoms, there may not be a chance of a cure by the time it is identified. This stage is characterized by pain, anorexia, early satiety, sleeplessness, and weight loss. Furthermore, a lower post-resection survival rate and unresectability are strongly indicated by back pain. Conversely, prompt detection of pancreatic cancer signs could improve the disease's early diagnosis. For example, 25% of patients may have symptoms consistent with upper abdominal disease up to six months before a diagnosis, and 15% of patients may seek medical attention more than six months before a diagnosis.

CHAPTER 3

AIM AND OBJECTIVES

3.1 AIM

- Pancreatic cancer is one of the major factors in our healthcare domain.
- There are a lot of patients who are actively present in the world.
- It is difficult to find the Pancreatic cancer. So this project can easily find out the attack.

3.2 OBJECTIVES

The goal is to develop a machine learning model for Pancreatic cancer, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithms.

3.3 SCOPE OF THE PROJECT

Here the scope of the project is that integration of patients' cancer with computer-based prediction could reduce errors and improve prediction outcomes. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment that can help to significantly improve the quality of Predicting pancreatic cancer.

CHAPTER 4

MATERIALS AND METHODOLOGY

4.1 WORKFLOW

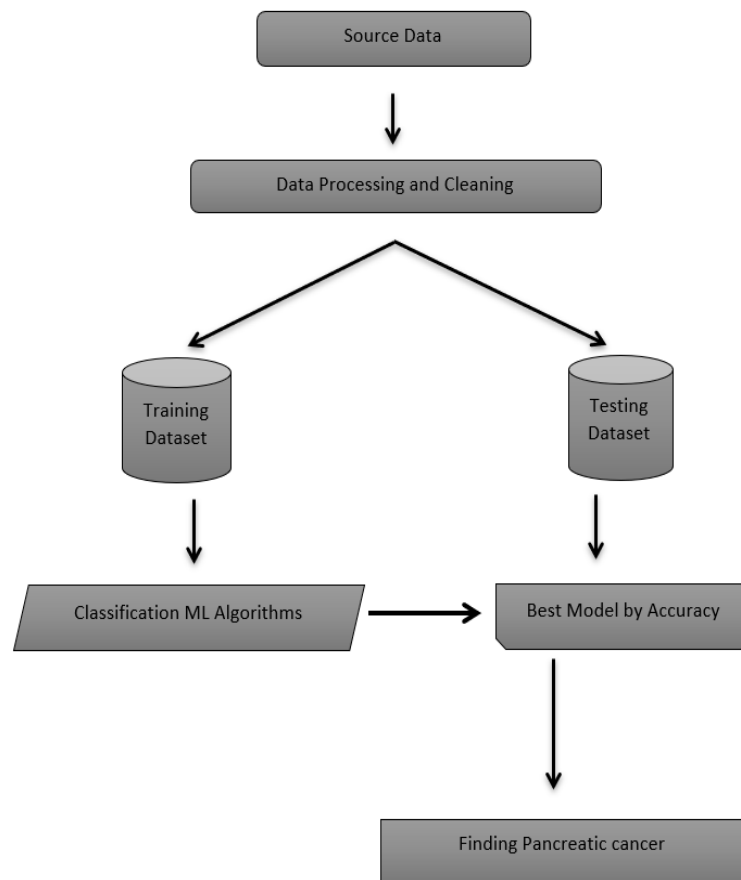


FIG4.1:WORK FLOW

4.2 DATA COLLECTION

The data set collected for predicting given data is split into a Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistics, Decision tree algorithms, and Support vector classifier (SVC) was applied to the Training set and based on the test result accuracy, Test set prediction was done.

4.2.1 KAGGLE DATASET

<https://www.kaggle.com/datasets/johnjdavisiv/urinary-biomarkers-for-pancreatic-cancer/data>

The key features are four urinary biomarkers: creatinine, LYVE1, REG1B, and TFF1.

- **Creatinine** is a protein that is often used as an indicator of kidney function.
- **YVLE1** is lymphatic vessel endothelial hyaluronan receptor 1, a protein that may play a role in tumor metastasis
- **REG1B** is a protein that may be associated with pancreas regeneration
- **TFF1** is trefoil factor 1, which may be related to regeneration and repair of the urinary tract

Age and **sex**, both included in the dataset, may also play a role in who gets pancreatic cancer. The dataset includes a few other biomarkers as well, but these were not measured in all patients (they were collected partly to measure how various blood biomarkers compared to urine biomarkers).

4.3 IMPORT LIBRARIES

```
# importing libraries
import pandas as pd
import numpy as np
# ignoring warnings
import warnings
warnings.filterwarnings('ignore')
```

sklearn

In Python, sklearn is a machine learning package that includes a lot of ML algorithms. Here, we are using some of its modules like `train_test_split`, `DecisionTreeClassifier`, or `Logistic Regression` and `accuracy_score`. (Dedov 2019)

NumPy

It is a numeric Python module that provides fast maths functions for calculations. It is used to read data in numpy arrays and for manipulation purposes. (Idris 2014)

Pandas

Used to read and write different files. Data manipulation can be done easily with data frames. (Stepanek 2020)

Matplotlib

Data visualization is a useful way to help identify the patterns from a given dataset. Data manipulation can be done easily with data frames. (Okura 2023)

4.4 PREPROCESSING THE URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING LABEL ENCODER

The data that was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion needs to be done.

This Python script performs data preprocessing tasks on a dataset related to pancreatic cancer research. It includes steps such as importing necessary libraries, reading the dataset, handling warnings, checking for null values and duplicated records, identifying categorical and numerical columns, and removing columns with significant null values.

After preprocessing, label encoding is applied to categorical features (excluding specific columns) to convert them into numerical representations. This prepares the dataset for further analysis or modeling.

4.5 VISUALIZING THE URINARY BIOMARKER FOR PANCREATIC CANCER USING MATPLOTLIB

Visualization plays a crucial role in understanding and interpreting machine learning models. By visually representing complex data and

model outputs, visualization techniques help data scientists and stakeholders gain insights into the behavior and performance of machine learning models. Through various visualization methods, such as plots, charts, and diagrams, key aspects of the model's performance, including accuracy, precision, recall, and feature importance, can be communicated effectively. Additionally, visualizations aid in identifying patterns, trends, and anomalies in the data, facilitating feature selection, model tuning, and decision-making processes

This Python script conducts data visualization tasks on a dataset related to pancreatic cancer research. It begins with importing necessary libraries, including pandas, numpy, seaborn, and matplotlib. After reading the dataset, null values are removed from specific columns to ensure data integrity. The data visualization section consists of six plots aimed at gaining insights into various aspects of the dataset:

A pie chart is used to check the balance of the dataset based on the 'diagnosis' column. A countplot is employed to compare the distribution of diagnoses across different sexes. A violin plot visualizes the distribution of ages among different diagnoses. Another countplot examines the distribution of sample origins across different diagnoses. A histogram plot depicts the distribution of creatinine levels among different diagnoses. Finally, a pair plot offers a comprehensive overview of the relationships between different numerical features in the dataset.

4.6 MODEL BUILT FOR URINARY MARKERS FOR PANCREATIC CANCER DATASET USING CATBOOST CLASSIFIER

The CatBoost Classifier, an advanced machine learning algorithm developed by Yandex, offers a comprehensive solution for handling categorical data directly within the modeling process. As part of the gradient boosting family, CatBoost stands out for its remarkable speed and accuracy in predictive modeling tasks. What sets CatBoost apart is its innovative approach to addressing the challenges of categorical variables, which are common in real-world datasets. The algorithm

employs a specialized algorithmic technique that effectively combats overfitting, ensuring robust performance even with complex datasets. One of Cat Boost's key strengths lies in its symmetric decision tree-based approach, augmented by ordered boosting, a method that enhances gradient estimation stability. Moreover, CatBoost prioritizes user convenience, providing seamless integration with popular data science frameworks and offering intuitive tools for model interpretation and analysis, including feature importance scores. Its versatility and effectiveness make CatBoost an ideal choice for a wide range of applications, from finance and marketing to medical diagnostics. In these domains, CatBoost excels in classification tasks, such as predicting customer churn, forecasting response to marketing campaigns, or identifying the onset of diseases. Overall, CatBoost represents a powerful tool in the data scientist's arsenal, capable of delivering high-performance predictive models while simplifying the challenges associated with categorical data handling. (Schumacher and LaBounty (2023))

The script that is provided entails importing data, cleaning it (by eliminating superfluous features and handling missing values), and using LabelEncoder to encode categorical variables in order to prepare a dataset for a genetic condition prediction. To avoid class imbalance problems, RandomOverSampler is used to balance the dataset. The balanced dataset is then used to train a CatBoostClassifier, and measures such as accuracy, confusion matrix, and classification report are used to assess the model's performance. This all-encompassing strategy guarantees a strong predictive model for hereditary illnesses.

4.7 MODEL BUILT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting

the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks. The following are the basic steps involved in performing the random forest algorithm, Pick N random records from the dataset, Build a decision tree based on these N records. In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote. (Phillips (2024))

The Python script reads a cancer diagnostic dataset, imports the required libraries for data manipulation and visualization, and does preliminary data cleaning by eliminating null values and superfluous columns. After encoding categorical variables using LabelEncoder, the data is divided into features (X) and target (y). The training and test sets of the dataset are then separated. Performance measures including accuracy, classification report, and confusion matrix are obtained by training and testing a RandomForestClassifier on the complete dataset. Ultimately, a function is constructed to show the confusion matrix graphically. The script efficiently expedites the process of preparing data and evaluating models for predictive analysis in the field of healthcare.

4.8 MODEL BUILT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING NAÏVE BAYES CLASSIFIER

The Naive Bayes classifier is a powerful probabilistic machine learning model used for classification tasks, particularly renowned for its simplicity and efficiency. Based on Bayes' Theorem, the classifier assumes independence among predictors within a dataset, a naive assumption that simplifies the computation, and despite its simplicity, often results in surprisingly effective classification performance. Naive Bayes is particularly advantageous when dealing with large datasets and performs well in scenarios where the independence assumption holds relatively true. It's commonly applied in text classification tasks such as spam detection, sentiment analysis, and document categorization. The algorithm's strength lies in its ability to quickly learn to use high-dimensional data and make predictions faster than more complex methods like neural networks. Naive Bayes classifiers also handle missing data gracefully and are less sensitive to irrelevant features, making them robust in varied applications. Furthermore, they are easy to implement, requiring a small amount of training data to estimate the necessary parameters. Naive Bayes provides a good baseline in many practical applications and is an excellent choice for projects where computational resources are limited, and real-time performance is critical. (Sabry (2023))

The code builds a Naive Bayes classifier for a medical diagnostic dataset. It starts by importing necessary libraries and preprocessing the data by removing irrelevant columns and handling missing values. Categorical variables are encoded using LabelEncoder. The dataset is then oversampled to address class imbalance. Next, it splits the data into training and test sets. The Naive Bayes classifier is trained on the training data and evaluated on the test set, yielding accuracy, confusion matrix, and classification report metrics. This streamlined approach ensures effective model development and evaluation for medical diagnosis.

4.9 DEPLOYMENT – DJANGO

Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and many options for free and paid-for support. (Dinder 2022)

Complete

Django follows the "Batteries included" philosophy and provides almost everything developers might want to do "out of the box". Because everything you need is part of the one "product", it all works seamlessly together, follows consistent design principles, and has extensive and up-to-date documentation.

Versatile

Django can be (and has been) used to build almost any type of website — from content management systems and wikis to social networks and news sites. It can work with any client-side framework and can deliver content in almost any format (including HTML, RSS feeds, JSON, XML, etc). The site you are currently reading is built with Django! Internally, while it provides choices for almost any functionality you might want (e.g. several popular databases, templating engines, etc.), it can also be extended to use other components if needed.

Secure

Django helps developers avoid many common security mistakes by providing a framework that has been engineered to "do the right things" to protect the website automatically. For example, Django provides a secure way to manage user accounts and passwords, avoiding common mistakes like putting session information in cookies where it is vulnerable

(instead cookies just contain a key, and the actual data is stored in the database) or directly storing passwords rather than a password hash. A password hash is a fixed-length value created by sending the password through a cryptographic hash function. Django can check if an entered password is correct by running it through the hash function and comparing the output to the stored hash value. However due to the "one-way" nature of the function, even if a stored hash value is compromised it is hard for an attacker to work out the original password. Django enables protection against many vulnerabilities by default, including SQL injection, cross-site scripting, cross-site request forgery, and clickjacking (see Website security for more details of such attacks).

Scalable

Django uses a component-based "shared-nothing" architecture (each part of the architecture is independent of the others, and can hence be replaced or changed if needed). Having a clear separation between the different parts means that it can scale for increased traffic by adding hardware at any level: caching servers, database servers, or application servers. Some of the busiest sites have successfully scaled Django to meet their demands (e.g. Instagram and Disqus, to name just two)

.

Maintainable

Django code is written using design principles and patterns that encourage the creation of maintainable and reusable code. In particular, it makes use of the Don't Repeat Yourself (DRY) principle so there is no unnecessary duplication, reducing the amount of code. Django also promotes the grouping of related functionality into reusable "applications" and, at a lower level, groups related code into modules (along the lines of the Model View Controller (MVC) pattern).

Portable

Django is written in Python, which runs on many platforms. That means that you are not tied to any particular server platform, and can run your applications on many flavors of Linux, Windows, and Mac OS X. Furthermore, Django is well-supported by many web hosting providers, who often provide specific infrastructure and documentation for hosting Django sites.

CHAPTER 5

RESULT AND DISCUSSION

5.1 RESULT

5.1.1 PREPROCESSING THE URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING LABEL ENCODER

```
df=pd.read_csv('Debernardi et al 2020 data.csv')
df.head(20)
# Total rows and columns
print('Total rows: {}\nTotal columns: {}'.format(df.shape[0],df.shape[1]))
# data types of each feature
df.info()
# Checking null values
df.isnull().sum()
# Checking duplicated records
print('duplicated records: ',df.duplicated().sum())
# categorical column vs numerical column
cat_columns=[feature for feature in df.columns if df[feature].dtypes=='O']
num_columns=[feature for feature in df.columns if df[feature].dtypes!='O']
print('Total categorical columns: ',len(cat_columns))
print('Total_numerical columns: ',len(num_columns))
# Unique features in numerical columns for feature in cat_columns:
print(feature+' :',len(df[feature].unique()))
# Numerical columns
print(num_columns)
# Describe function for numerical columns
df.describe()
# Correlation between the numerical columns
df.corr()
```

This Python script preprocesses and analyzes data from a medical

dataset. To improve readability, it first imports the required libraries, including Pandas and NumPy, and turns off warnings. Basic exploratory analysis is performed on the dataset when it is read from a CSV file. This includes showing the first 20 rows, confirming that there are exactly 20 rows and columns, and looking at data types, null values, and duplicate entries. Numerical and categorical columns are distinguished, and each categorical attribute has a distinct value shown. Removing null values involves removing particular columns that have a large amount of missing data. Categorical features are converted into numerical representations through the use of label encoding. The preprocessed dataset is finally shown, highlighting the modifications.

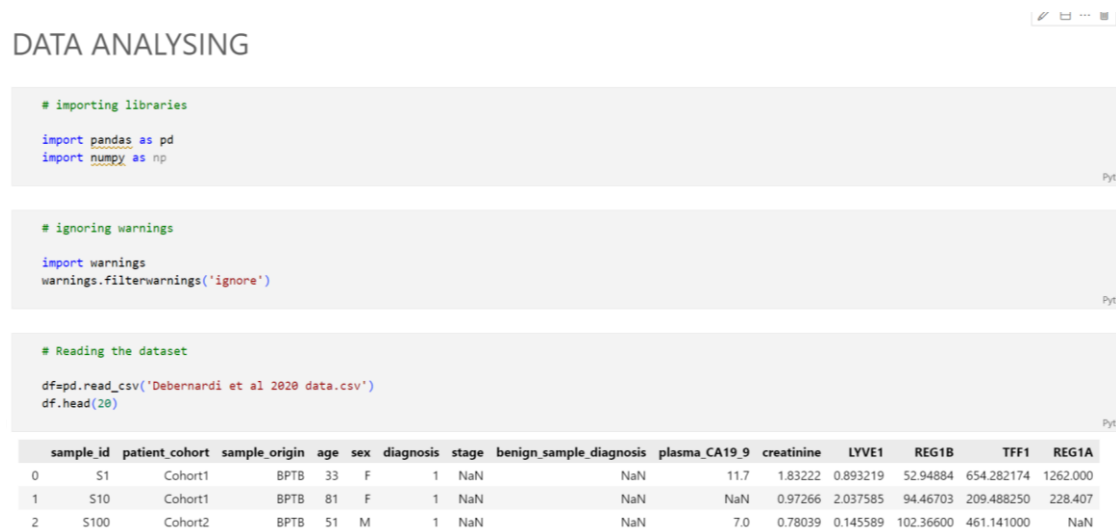


FIG.5.1 PREPROCESSING THE URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET

5.1.2 VISUALIZING THE URINARY BIOMARKER FOR PANCREATIC CANCER USING MATPLOTLIB

```
pieplot=df['diagnosis'].value_counts()  
ax=pieplot.plot.pie(figsize=(12,6), autopct='%1.2f%%',  
explode=[0.01,0.05,0.1], shadow=True)
```

The code snippet provided uses the pandas and matplotlib libraries in Python to create a pie chart visualization of the distribution of values in the 'diagnosis' column of a data frame called df.

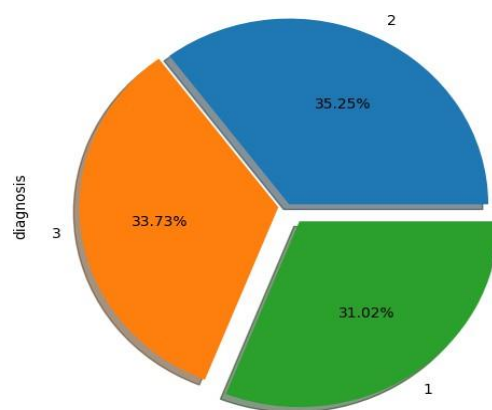


FIG5.2: PIE PLOT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER

`df['diagnosis'].value_counts()`: This function counts the occurrences of each unique value in the 'diagnosis' column of the DataFrame df. It returns a Series where the index contains the unique values and the corresponding values are the counts of these occurrences.

`pieplot.plot.pie(figsize=(12,6),autopct='%1.2f%%',explode=[0.01,0.05,0.1], shadow=True)`, `figsize=(12,6)`: Sets the size of the figure (width, height) in inches.

`autopct='%1.2f%%'`: Formats the percentage labels inside the pie chart to two decimal places. This shows each slice's percentage of the total.

`explode=[0.01,0.05,0.1]`: This parameter offsets slices from the pie chart to emphasize them. The offsets are given as fractions of the pie radius, and the list should have the same number of elements as there are slices

in the pie. Each value corresponds to how much a particular slice is exploded from the center.

shadow=True: Adds a shadow beneath the pie chart, giving it a three-dimensional appearance.

```
sns.countplot(x=df['diagnosis'],hue=df['sex'])
```

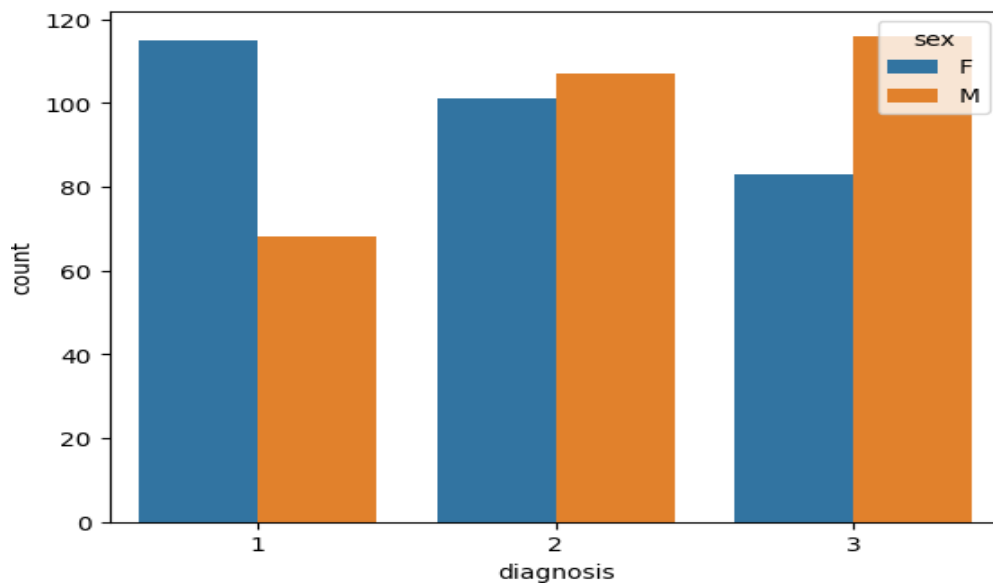


FIG5.3COMPARING DIAGNOSIS WITH SEX IN URINARY BIOMARKERS FOR PANCREATIC CANCER

sns.countplot(): This function from the seaborn library creates a countplot.

x=df['diagnosis']: Specifies the data to be plotted on the x-axis. In this case, it's the 'diagnosis' column of the DataFrame df, which likely contains categorical data.

hue=df['sex']: This parameter adds an additional categorical separation. It essentially creates separate bars for each category in 'diagnosis', and each of these bars is further divided into sub-bars based on the 'sex' column. So, for each category in 'diagnosis', you'll have bars representing the count of observations for both male and female, side by side.

```
sns.violinplot(x=df['diagnosis'],y=df['age'])
```

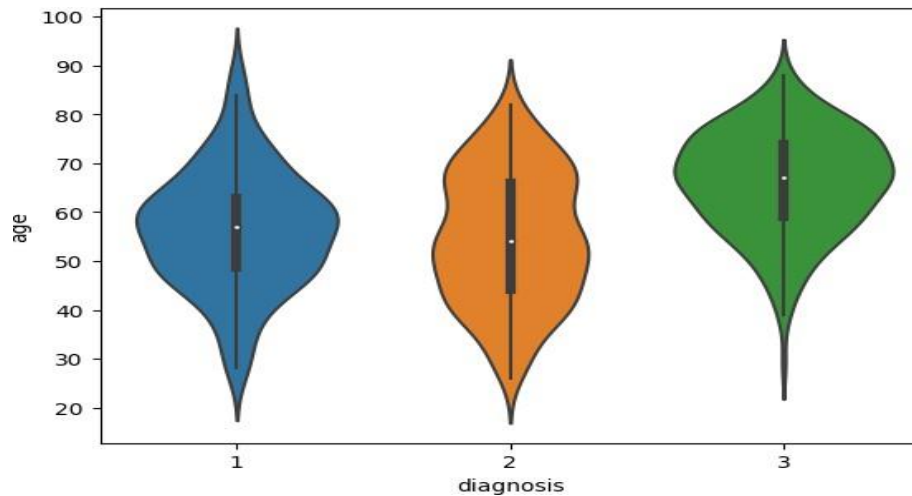


FIG.5.4: COMPARING DIAGNOSIS WITH AGE IN URINARY BIOMARKERS FOR PANCREATIC CANCER

`sns.violinplot()`: This function from the seaborn library creates a violin plot.
`x=df['diagnosis']`: Specifies the categorical data to be plotted on the x-axis. In this case, it's the 'diagnosis' column of the DataFrame `df`, which likely contains different categorical diagnoses (such as types of diseases or conditions).
`y=df['age']`: Specifies the numeric data to be plotted on the y-axis. Here, it is the 'age' column, representing the age of the subjects or patients.

```
sns.countplot(x=df['sample_origin'],hue=df['diagnosis'])
```

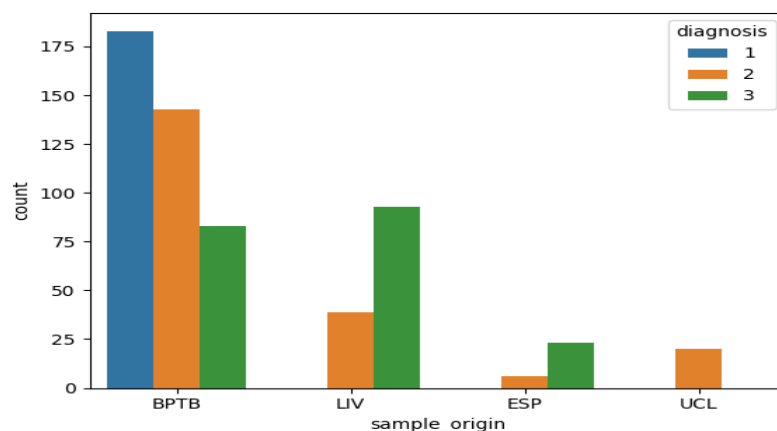
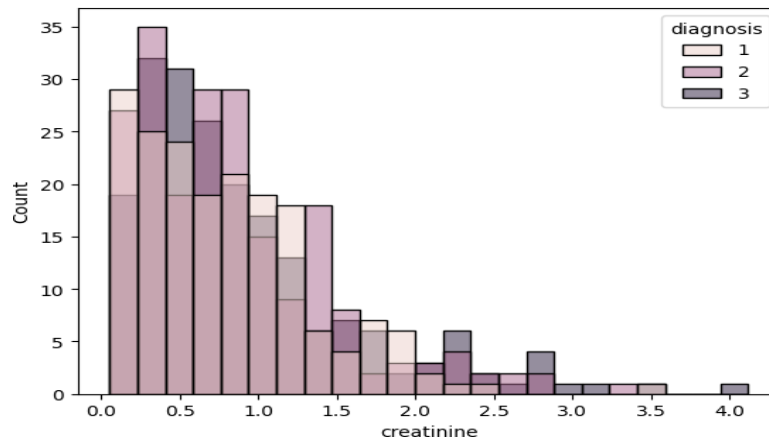


FIG5.5: COMPARING SAMPLE ORIGIN VS DIAGNOSIS IN URINARY BIOMARKERS FOR PANCREATIC CANCER

`sns.countplot()`: This function creates a bar plot that shows the count of observations in each categorical bin using `bars.x=df['sample_origin']`: This sets the 'sample_origin' column of the DataFrame `df` as the data for the x-axis. The 'sample_origin' column likely contains categorical data indicating where each sample came from (such as different hospitals, clinics, or other medical facilities). `hue=df['diagnosis']`: This parameter adds a layer of categorical separation within each primary category on the x-axis. By setting `hue` to 'diagnosis', the countplot will not only show the total counts of 'sample_origin' but will also break down these counts by different types of diagnoses

`sns.histplot(x=df['creatinine'],hue=df['diagnosis'])`



**FIG5.6: COMPARING CREATININE VS DIAGNOSIS IN
URINARY BIOMARKERS FOR PANCREATIC CANCER**

```
sns.pairplot(df)
```

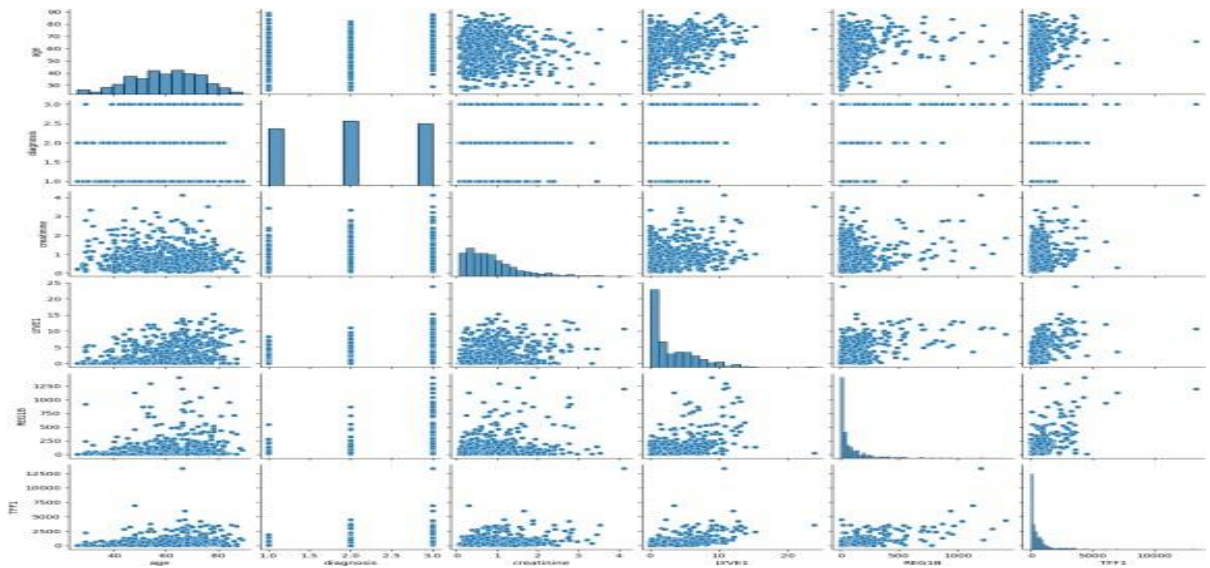


FIG. 5.7 PAIR PLOT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER

5.1.3 MODEL BUILT FOR URINARY MARKERS FOR PANCREATIC CANCER DATASET USING CATBOOST CLASSIFIER

```
# Transform the datasets string value into integers values or float value(Label
encoder)
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
col = ["Patient Age", "Genes in mother's side", "White Blood cell count (thousand
per microliter)", "Paternal gene", "Blood cell count (mcL)", "Respiratory Rate
(breaths/min)", "Heart Rate (rates/min)", "Parental consent", "Follow-up", "Gender",
'Folic acid details (peri-conceptional)', 'H/O serious maternal illness', 'H/O
substance abuse', 'Assisted conception IVF/ART', 'Blood test result', 'Genetic
Disorder', 'Inherited from father', 'Maternal gene']
for i in col:
df[i] = le.fit_transform(df[i]).astype(int)
# Check the data relationships
df.head()
```

```

# Check the unique values in specific columns
df['Genetic Disorder'].unique()
x1 = df.drop(labels='Genetic Disorder',axis=1)
y1 = df.loc[:, 'Genetic Disorder']
import imblearn
from imblearn.over_sampling import RandomOverSampler
from collections import Counter
ros =RandomOverSampler(random_state=1)
x,y=ros.fit_resample(x1,y1)
print("OUR DATASET COUNT      :", Counter(y1))
print("OVER SAMPLING DATA COUNT :", Counter(y))
from sklearn.model_selection import train_test_split
x_train,x_test, y_train, y_test = train_test_split(x,y, test_size=0.20, shuffle=False)
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import AdaBoostClassifier
from sklearn.impute import SimpleImputer
from sklearn.pipeline import make_pipeline
from catboost import CatBoostClassifier # Algorithms ok
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.tree import ExtraTreeClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.ensemble import HistGradientBoostingClassifier
#clf = make_pipeline(SimpleImputer(strategy='mean'), ExtraTreesClassifier())
clf = CatBoostClassifier()
clf.fit(x_train, y_train)
predictLR =clf.predict(x_test)

```

```

from sklearn.metrics import accuracy_score

accuracy_score = accuracy_score(y_test, predictLR)* 100
print("Accuracy Score\n\n", accuracy_score)

```

Accuracy Score

86.14916286149162

FIG5.8:ACCURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING CATBOOST CLASSIFIER

Data Preparation: To generate the feature set x1, remove the column "Genetic Disorder" from your DataFrame. To generate the target variable y1, all you have to do is select the "Genetic Disorder"column.

Managing Class Imbalance: To handle class imbalance, you are oversampling your data using RandomOverSampler. The count of each class in the original dataset is printed by Counter(y1). After oversampling, Counter(y) prints the count for each class. Train-Test Split: Training and testing sets are created from the oversampled data (x, y). The features and labels for the training set are denoted by x_train and y_train, whereas those for the testing set are denoted by x_test and y_test. Test_size=0.20 indicates that 20% of the data will be utilized for testing; shuffle=False indicates that there will be no shuffle of the data prior to splitting.

5.1.4 MODEL BUILT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING RANDOM FOREST

Transform the datasets string value into integers values or float value(Label encoder)

```

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
col = ["sample_origin", "sex", 'stage', 'patient_cohort']
for i in col:
    data[i] = le.fit_transform(data[i]).astype(int)

```

```
# Split the datasets into training process X is (independent features, independent features)
```

```
x1 = data.drop(labels='diagnosis',axis=1)
```

```
y1 = data.loc[:, 'diagnosis']
```

```
x1.info()
```

```
# Balance the datasets
```

```
import imblearn
```

```
from imblearn.over_sampling import RandomOverSampler
```

```
from collections import Counter
```

```
ros =RandomOverSampler(random_state=1)
```

```
x,y=ros.fit_resample(x1,y1)
```

```
print("OUR DATASET COUNT      :", Counter(y1))
```

```
print("OVER SAMPLING DATA COUNT :", Counter(y))
```

```
x.info()
```

```
# Split the datasets into training and testing
```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test, y_train, y_test = train_test_split(x,y, test_size=0.20,  
random_state=42)
```

```
# Import the model build algorithm
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
clf = RandomForestClassifier()
```

```
clf.fit(x_train, y_train)
```

```
predictRFC =clf.predict(x_test)
```

```
# Check the accuracy score of this model
```

```
from sklearn.metrics import accuracy_score
```

```
accuracy_score = accuracy_score(y_test, predictRFC)* 100
```

```
print("Accuracy Score is ", accuracy_score)
```

```
# Check the accuracy score of this model
from sklearn.metrics import accuracy_score

accuracy_score = accuracy_score(y_test, predictRFC)*100
print("Accuracy Score is ", accuracy_score)
```

Accuracy Score is 92.80000000000001

FIG5.9.ACUURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING RANDOM FOREST

Test Size = 0.20 random_state = 42

Data Preparation: To generate the feature set x1, you are removing the column "diagnosis" from your DataFrame. To build the target variable y1, you are just choosing the column labeled "diagnosis." Examining the Initial Data: To examine the data, you have shown the first 20 rows of the feature set x1.

Balancing the Dataset: By oversampling the minority class, RandomOverSampler is used to correct the class imbalance. To demonstrate the effect of balancing, you are printing the count of each class both before and after oversampling.

Verifying Balanced Data: To examine the structure of the oversampled feature set x, you have run info().

5.1.5 MODEL BUILT FOR URINARY BIOMARKERS FOR PANCREATIC CANCER DATASET USING NAÏVE BAYES CLASSIFIER

```
# Split the datasets into training process X is (independent features, independent
features)
x1 = data.drop(labels='diagnosis',axis=1)
y1 = data.loc[:, 'diagnosis']
import imblearn
from imblearn.over_sampling import RandomOverSampler
from collections import Counter
ros =RandomOverSampler()
x,y=ros.fit_resample(x1,y1)
print("OUR DATASET COUNT      :", Counter(y1))
print("OVER SAMPLING DATA COUNT :", Counter(y))
from sklearn.model_selection import train_test_split
x_train,x_test, y_train, y_test = train_test_split(x,y, test_size=0.20, shuffle=False)
from sklearn.naive_bayes import GaussianNB
clf = GaussianNB()
clf.fit(x_train, y_train)
predictLR =clf.predict(x_test)
from sklearn.metrics import accuracy_score
accuracy_score = accuracy_score(y_test, predictLR)* 100
print("Accuracy Score\n\n", accuracy_score)
```

```
from sklearn.metrics import accuracy_score
accuracy_score = accuracy_score(y_test, predictLR)*100
print("Accuracy Score\n\n", accuracy_score)
```

Accuracy Score

99.2

FIG5.10:ACCURACY OF URINARY BIOMARKERS FOR PANCREATIC CANCER USING NAIVEBAYES CLASSIFIER

5.1.6 DEPLOYMENT

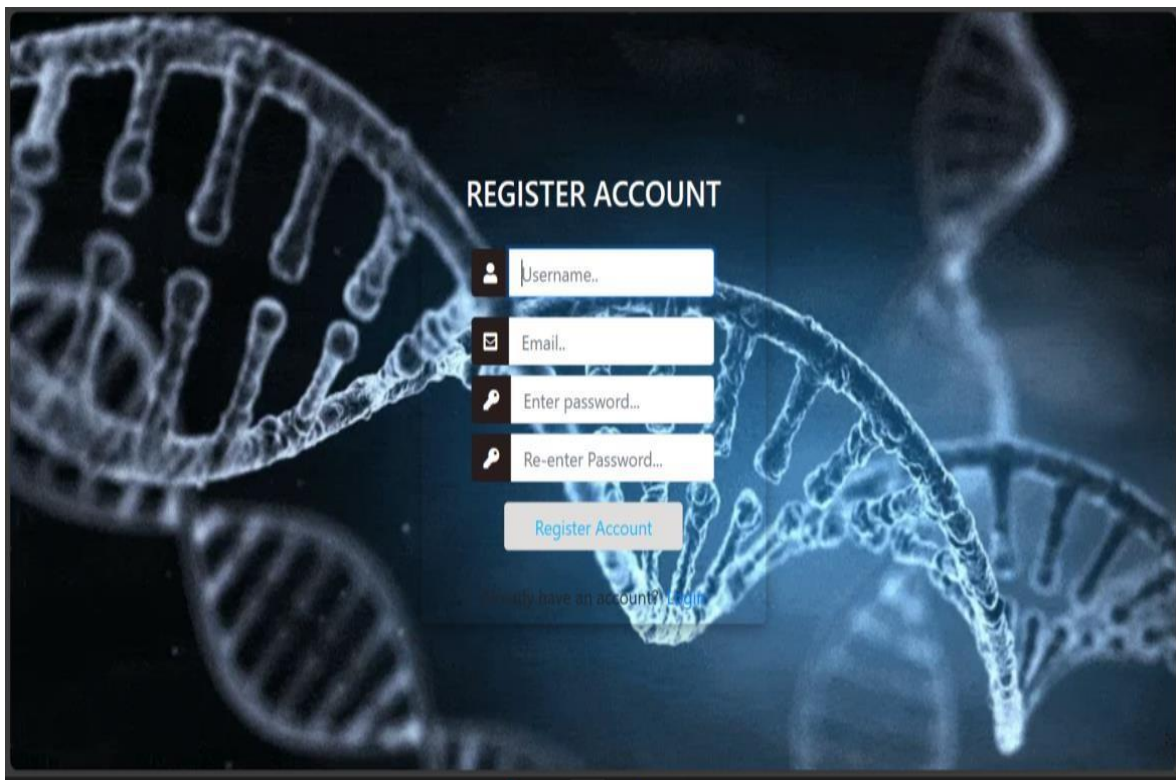


FIG5.11:URINARYBIOMARKERS FOR PANCREATIC CANCER PREDICTION LOGING PAGE

P-Cancer Home Domain Overview Problem statement User Information Model User_Database Model Database Logout

Model Analysis

PATIENT_AGE
Please enter your age

PLASMA_CA19_9
Please enter plasma_CA19_9 value

STAGE
Please enter your Stage

CREATININE
Please enter blood_cell_count value

LYVE1
Please enter LYVE1 value

FIG5.12: URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION WEBPAGE(A)

Please enter blood_cell_count value

LYVE1
Please enter LYVE1 value

REG1B
Please enter REG1B value

TFF1
Please enter TFF1 value

REG1A
Please enter TFF1 value

PATIENT_COHORT
Cohort2

GENDER
Female

SAMPLE_ORIGIN
BPTB

Submit

FIG5.13: URINARY BIOMARKERS FOR PANCREATIC CANCER PREDICTION WEBPAGE(B)

5.2 DISCUSSION

Developing a machine learning model for pancreatic cancer categorization is the suggested approach. The procedure starts with data gathering, where historical information about pancreatic cancer is gathered. One type of data mining is widely utilized method for handling massive amounts of data in the healthcare industry. If detected early enough, pancreatic cancer can be treated to prevent death.

These days, machine learning is largely utilized in the medical field because it lowers manual labor and improves models, reducing errors and perhaps saving lives.

The dataset is properly variable identified, meaning that both the independent and dependent variables are located before the data analysis is completed. Subsequently, appropriate machine learning algorithms are implemented on the dataset containing the data pattern.

CHAPTER 6

SUMMARY AND CONCLUSION

The exploration of urinary biomarkers for pancreatic cancer prediction is a critical endeavor in advancing early detection and treatment strategies for this disease. By leveraging machine learning algorithms and statistical techniques, researchers aim to identify specific biomarkers present in urine samples that can serve as reliable indicators of pancreatic cancer presence or progression.

Initial steps typically involve data collection, including urine samples from both pancreatic cancer patients and healthy individuals for comparison. Various analytical methods, such as mass spectrometry or immunoassays, may be employed to identify potential biomarkers present in the urine samples. Once the data is collected and biomarkers are identified, machine learning algorithms are trained using this data to develop predictive models. These models aim to classify urine samples as either indicative of pancreatic cancer or not based on the presence or absence of specific biomarkers.

Evaluation of these models involves rigorous testing using independent datasets to assess their accuracy, sensitivity, specificity, and other performance metrics. Additionally, validation studies are often conducted to confirm the reliability and reproducibility of the identified biomarkers and predictive models across different populations and sample cohorts.

The analytical process started with data cleaning and processing, missing value, exploratory analysis, and finally model building and evaluation. The best accuracy on the public test set of higher accuracy score algorithms will be found out. The founded one is used in the application which can help to find the Pancreatic cancer of the patient.

REFERENCES

1. Aalami, A. H., Abdeahad, H., Mesgari, M., Sathyapalan, T., & Sahebkar, A. (2021). Urinary Angiogenin as a Marker for Bladder Cancer: A Meta-Analysis. *BioMed research international*, 2021, 5557309. <https://doi.org/10.1155/2021/5557309>
2. Advances in biomarkers and techniques for pancreatic cancer diagnosis. *Cancer Cell Int.* 2022;22(1):220. doi: 10.1186/s12935-02202640-9.
3. Cheema AR, O'Reilly EM. Management of Metastatic Pancreatic Adenocarcinoma. *Surg Clin North Am.* 2016 Dec;96(6):1391-1414. PMID: 27865284
4. Cui, Y., Shu, X. O., Li, H. L., Yang, G., Wen, W., Gao, Y. T., Cai, Q., Rothman, N., Yin, H. Y., Lan, Q., Xiang, Y. B., & Zheng, W. (2017). Prospective study of urinary prostaglandin E2 metabolite and pancreatic cancer risk. *International journal of cancer*, 141(12), 2423–2429. <https://doi.org/10.1002/ijc.31007>
5. De Matteis, S., Bonafè, M., & Giudetti, A. M. (2021). Urinary Metabolic Biomarkers in Cancer Patients: An Overview. *Methods in molecular biology* (Clifton, N.J.), 2292, 203–212. https://doi.org/10.1007/978-1-0716-1354-2_18
6. de Vries, E. G. E., Rüschoff, J., Lolkema, M., Tabernero, J., Gianni, L., Voest, E., de Groot, D. J. A., Castellano, D., Erb, G., Naab, J., Donica, M., Deurloo, R., van der Heijden, M. S., & Viale, G. (2023). Phase II study (KAMELEON) of single-agent T-DM1 in patients with HER2-positive advanced urothelial bladder cancer or pancreatic cancer/cholangiocarcinoma. *Cancer medicine*, 12(11), 12071–12083. <https://doi.org/10.1002/cam4.5893>

7. Debernardi, S., Blyuss, O., Rycyk, D., Srivastava, K., Jeon, C. Y., Cai, H., Cai, Q., Shu, X. O., & Crnogorac-Jurcevic, T. (2023). Urine biomarkers enable pancreatic cancer detection up to 2 years before diagnosis. *International journal of cancer*, 152(4), 769–780. <https://doi.org/10.1002/ijc.34287>

8. Debernardi, S., Massat, N. J., Radon, T. P., Sangaralingam, A., Banissi, A., Ennis, D. P., Dowe, T., Chelala, C., Pereira, S. P., Kocher, H. M., Young, B. D., Bond-Smith, G., Hutchins, R., & Crnogorac-Jurcevic, T. (2015). Noninvasive urinary miRNA biomarkers for early detection of pancreatic adenocarcinoma. *American journal of cancer research*, 5(11), 3455–3466.

9. DiMagno EP. Pancreatic cancer: clinical presentation, pitfalls and early clues. *Ann Oncol.* 1999;10 Suppl 4:140-2. PMID: 10436807

10. Debernardi, S., O'Brien, H., Algahmdi, A. S., Malats, N., Stewart, G. D., Plješa-Ercegovac, M., Costello, E., Greenhalf, W., Saad, A., Roberts, R., Ney, A., Pereira, S. P., Kocher, H. M., Duffy, S., Blyuss, O., & Crnogorac-Jurcevic, T. (2020). A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case-control study. *PLoS medicine*, 17(12), e1003489. <https://doi.org/10.1371/journal.pmed.1003489>

11. González-Gómez R, Pazo-Cid RA, Sarría L, Morcillo MÁ, Schuhmacher AJ: Diagnosis of Pancreatic Ductal Adenocarcinoma by Immuno-Positron Emission Tomography. *J Clin Med.* 2021; 10:1151

12. Hipperson, L., Hadden, W. J., Nahm, C. B., Gill, A. J., Samra, J. S., Dona, A., Mittal, A., & Sahni, S. (2021). Urinary metabolite prognostic biomarker panel for pancreatic ductal adenocarcinomas. *Biochimica et biophysica acta. General subjects*, 1865(11), 129966. <https://doi.org/10.1016/j.bbagen.2021.129966>

13. Husi, H., Fernandes, M., Skipworth, R. J., Miller, J., Cronshaw, A. D., Fearon, K. C. H., & Ross, J. A. (2019). Identification of diagnostic upper gastrointestinal cancer tissue type-specific urinary biomarkers. *Biomedical reports*, 10(3), 165–174. <https://doi.org/10.3892/br.2019.1190>

14. Bertin M, Bourgon R, Coil K et al: Abstract PO-007: Plasma-based detection of pancreatic cancer: A multiomics approach. *Cancer Res* 2021, 81(22_Supplement):PO-007-PO-007.

15. Karar, M. E., El-Fishawy, N., & Radad, M. (2023). Automated classification of urine biomarkers to diagnose pancreatic cancer using 1-D convolutional neural networks. *Journal of biological engineering*, 17(1), 28. <https://doi.org/10.1186/s13036-023-00340-0>

16. Karar, M. E., Alotaibi, B., & Alotaibi, M. (2022). Intelligent Medical IoT-Enabled Automated Microscopic Image Diagnosis of Acute Blood Cancers. *Sensors (Basel, Switzerland)*, 22(6), 2348. <https://doi.org/10.3390/s22062348>

17. Nissinen, S. I., Venäläinen, M., Kumpulainen, P., Roine, A., Häkkinen, M. R., Vepsäläinen, J., Oksala, N., & Rantanen, T. (2021). Discrimination between Pancreatic Cancer, Pancreatitis and Healthy Controls Using Urinary Polyamine Panel. *Cancer control : journal of the Moffitt Cancer Center*, 28, 10732748211039762. <https://doi.org/10.1177/10732748211039762>

18. O'Neill, R. S., & Stoita, A. (2021). Biomarkers in the diagnosis of pancreatic cancer: Are we closer to finding the golden ticket?. *World journal of gastroenterology*, 27(26), 4045–4087. <https://doi.org/10.3748/wjg.v27.i26.4045>

19. Sahni, S., Pandya, A. R., Hadden, W. J., Nahm, C. B., Maloney, S., Cook, V., Toft, J. A., Wilkinson-White, L., Gill, A. J., Samra, J. S., Dona, A., & Mittal, A. (2021). A unique urinary metabolomic signature for the detection of pancreatic ductal adenocarcinoma. *International journal of cancer*, 148(6), 1508–1518. <https://doi.org/10.1002/ijc.33368>
20. Wang, Y., Liu, K., Ma, Q., Tan, Y., Du, W., Lv, Y., Tian, Y., & Wang, H. (2019). Pancreatic cancer biomarker detection by two support vector strategies for recursive feature elimination. *Biomarkers in medicine*, 13(2), 105–121. <https://doi.org/10.2217/bmm-2018-0273>
21. Yip-Schneider, M. T., Soufi, M., Carr, R. A., Flick, K. F., Wu, H., Colgate, C. L., & Schmidt, C. M. (2020). Performance of candidate urinary biomarkers for pancreatic cancer - Correlation with pancreatic cyst malignant progression?. *American journal of surgery*, 219(3), 492–495. <https://doi.org/10.1016/j.amjsurg.2019.09.013>
22. Yip-Schneider, M. T., Soufi, M., Carr, R. A., Flick, K. F., Wu, H., & Schmidt, C. M. (2020). Discussion on: Performance of candidate urinary biomarkers for pancreatic cancer - Correlation with pancreatic cyst malignant progression?. *American journal of surgery*, 219(3), 496. <https://doi.org/10.1016/j.amjsurg.2020.02.014>