Electronic Health Record (EHR) System for Pancreatic Cancer and Inflammatory Arthritis

Department of Biostatistics, Georgetown University
Washington, DC
2020.12.5

Introduction

Pancreatic cancer, a disease in which malignant (cancer) cells form in the tissues of the pancreas, one of the deadliest cancers worldwide, has a five-year survival rate of less than 6%, the lowest percentage for cancers in the period 2007–2013[1]. Signs and symptoms of pancreatic cancer include jaundice, pain, weight loss and pain in the upper or middle abdomen and back. Unlike other types of cancer, few risk factors have been attributed to pancreatic cancer[2]. Under this situation, the potential role of inflammation, a reaction to an irritant, causing redness, swelling and pain, in the development and growth of cancer was initially described in 1863 by Virchow[3], who observed that inflammatory cells infiltrate tumors. Therefore, it is possible that systemic and local chronic inflammation might enhance the risk of pancreatic cancer. Meanwhile, Inflammatory arthritis (IA) is joint inflammation caused by an overactive immune system. It usually affects many joints throughout the body at the same time, which is often treated with a combination of medications that relieve swelling and pain along with others, such as steroids or immunosuppressive drugs, that regulate the immune system. According to Patel, R., new onset PDAC in patients treated with approved biologics for the treatment of IA has been described[4]. Based on these situations that described above, it might be a meaningful step to explore the relationship between pancreatic cancer patients and inflammatory arthritis patients.

This study aims to build a dynamic Electronic Health Record (EHR) System which is able to keep updating the records for pancreatic cancer and inflammatory arthritis. By using this system, it is possible to explore more about the connection of pancreatic cancer and inflammatory arthritis along with building a diagnostic model for these two diseases.

System Description

In our Electronic Health Record (EHR) system, we created a relational database to store the medical data for the patients. The machine learning algorithms are used for predicting the risk of the disease.

Functional requirement

Electronic Health Record (EHR) system is expected to include a dynamic updating database with the functional requirement including enquiries and adding both patient and exam information. After the analysis, our EHR system mainly includes the following several functions:

(1) Patient Information

Enquiries: Query the personal information for a specific patient.

Adding: New patient registration.

(2) Exam Information

Enquiries: Query the exam information for a specific patient.

Sorting: Sorting the exam information by date.

Adding: New exam information update.

Functional dependencies

Enquiries refers to querying the information for a specific patient while adding means adding the new information for a new registration patient. The translation to relational schema includes Patient (Patient_id, First_Name, Last_Name, Gender, Birth_Date, Race, Education, Email, Contact), Pancreatic_Cancer (Patient_id, PPI_Use, Pancreatic_Cancer, Exam_date) and IA_Test_Record (Patient_id, Duration, DMARDs, PDQ_score, Inflammatory_Arthritis, Exam_date, BMI, SBP, DBP, Smoking_Status, Diabetes). The functional dependencies that optimize the translation relational schema for this database is shown in Fig 1. Patient_id is a Foreign key from relation Patient, attribute Patient_id.

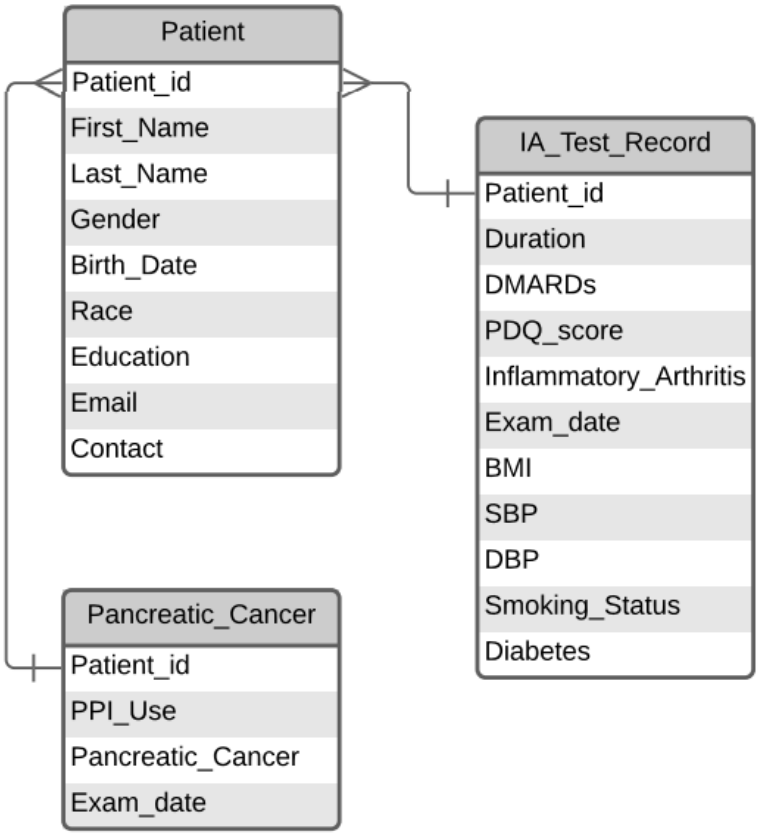

Fig 1. The functional dependencies for the database.

Database Structure

According to the overall database structure, the basic table structure of the Electronic Health Record (EHR) system is designed, and the corresponding target is defined as Table 1-3. And the corresponding SQL code could be found in supplements.

Table 1. Patient Information

Attribute	**Type**	**Length**	**Can be null or not**	**Explain**
Patient_id	varchar	255	primary key	Identity number of patients
First_Name	varchar	255	not null	First Name
Last_Name	int	11	not null	Last Name
Gender	varchar	255	not null	Gender
Birth_Date	date	0	not null	Birth Date
Race	varchar	255	not null	Race

Education	varchar	255	not null	Education
Email	varchar	255	not null	Email of patient
Contact	varchar	255	not null	Contact of patient

Table 2. Pancreatic Cancer Information

Attribute	**Type**	**Length**	**Can be null or not**	**Explain**
Patient_id	varchar	255	primary key	Identity number of patients
PPI_Use	varchar	255	not null	Proton Pump-Inhibitor Use (Yes=1, No=0)
Pancreatic_Cancer	int	11	not null	Pancreatic Cancer status (Yes=1, No=0)
Exam_date	date	0	not null	Date of the exam

Table 3. Inflammatory Arthritis Test Record Information

Attribute	**Type**	**Length**	**Can be null or not**	**Explain**
Patient_id	varchar	255	primary key	Identity number of patients
Duration	float	0	not null	Disease duration
DMARDs	int	11	not null	Disease modifying anti-rheumatic drugs
PDQ_score	int	11	not null	Pain DETECT Questionnaire score
Inflammatory_Arthritis	int	11	not null	Inflammatory Arthritis Status (Yes=1, No=0)
Exam_date	date	0	not null	Date of the exam
BMI	float	0	not null	Body Mass Index
SBP	float	0	not null	Systolic Blood Pressure
DBP	float	0	not null	Diastolic Blood Pressure
Smoking_Status	int	11	not null	Smoking_Status(Yes=1, No=0)
Diabetes	int	11	not null	Diabetes Status(Yes=1, No=0)

System Implementation

Database Implementation

First create the database of the Electronic Health Record (EHR) system, and then create the six basic tables in the database structure. The specific SQL code is shown as follows:

1. Create the database

```
CREATE DATABASE ehr;
```

2. Create the table of Patient Information

```
DROP TABLE IF EXISTS `patient`;
CREATE TABLE `patient` (
  `Patient_id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Identity number of patient',
  `First_Name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'First Name',
  `Last_Name` int NOT NULL COMMENT 'Last Name',
  `Gender` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Gender',
  `Birth_Date` date NOT NULL COMMENT 'Birth Date',
  `Race` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Race',
  `Education` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Education',
  `Email` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Email of patient',
  `Contact` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Contact of patient',
  PRIMARY KEY (`Patient_id`) USING BTREE
) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4_0900_ai_ci ROW_FORMAT = Dynamic;
```

3. Create the table of Pancreatic_Cancer information

```
DROP TABLE IF EXISTS `pancreatic_cancer`;
CREATE TABLE `pancreatic_cancer` (
  `Patient_id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Identity number of patient',
  `PPI_Use` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL COMMENT 'Proton Pump-Inhibitor Use(Yes=1, No=0)',
  `Pancreatic_Cancer` int NOT NULL COMMENT 'Pancreatic Cancer status\r\n(Yes=1, No=0)',
```

```
  `Exam_date` date NOT NULL COMMENT 'Date of the exam',
  PRIMARY KEY (`Patient_id`) USING BTREE,
  CONSTRAINT `Patient_id` FOREIGN KEY (`Patient_id`) REFERENCES `patient`
(`Patient_id`) ON DELETE RESTRICT ON UPDATE RESTRICT
) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4_0900_ai_ci
ROW_FORMAT = Dynamic;
```

4. Create the table of IA_Test_Record information

```
DROP TABLE IF EXISTS `ia_test_record`;
CREATE TABLE `ia_test_record` (
  `Patient_id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT
NULL COMMENT 'Identity number of patient',
  `Duration` float NOT NULL COMMENT 'Disease duration',
  `DMARDs` int NOT NULL COMMENT 'Disease modifying anti-rheumatic drugs',
  `PDQ_score` int NOT NULL COMMENT 'Pain DETECT Questionnaire score',
  `Inflammatory_Arthritis` int NOT NULL COMMENT 'Inflammatory Arthritis Status (Yes=1,
No=0)',
  `Exam_date` date NOT NULL COMMENT 'Date of the exam',
  `BMI` float NOT NULL COMMENT 'Body Mass Index',
  `SBP` float NOT NULL COMMENT 'Systolic Blood Pressure',
  `DBP` float NOT NULL COMMENT 'Diastolic Blood Pressure',
  `Smoking_Status` int NOT NULL COMMENT 'Smoking_Status(Yes=1, No=0)',
  `Diabetes` int NOT NULL COMMENT 'Diabetes Status(Yes=1, No=0)',
  PRIMARY KEY (`Patient_id`) USING BTREE,
  CONSTRAINT `Pid` FOREIGN KEY (`Patient_id`) REFERENCES `patient` (`Patient_id`) ON
DELETE RESTRICT ON UPDATE RESTRICT
) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4_0900_ai_ci
ROW_FORMAT = Dynamic;
```

Data Loading

1. Insert the data into table Patient

```
INSERT INTO `patient` VALUES ('10001', 'Malyn', 0, 'F', '1997-05-06', 'Asian', 'Associate
degre', 'JaneSmithChicago@ESP.com', '14530192012');
INSERT INTO `patient` VALUES ('10002', 'Jonet', 0, 'F', '1995-06-10', 'White', 'Bachelor\'s
degree', 'MayurDikShit@example.com', '10018585115');
INSERT INTO `patient` VALUES ('10003', 'Alise', 0, 'F', '1992-05-09', 'Black or African
American', 'Associate degre', 'Google@example.com', '17417413005');
INSERT INTO `patient` VALUES ('10004', 'Salove', 0, 'F', '1998-10-11', 'Hispanic or Latino',
'High school diploma or equivalent', 'party@college.edu', '12054232319');
INSERT INTO `patient` VALUES ('10005', 'Acelina', 0, 'F', '2000-05-30', 'American Indian or
Alaska Native', 'High school diploma or equivalent', 'ironman@timgarage.com',
'13469515966');
```

.....
INSERT INTO `patient` VALUES ('10019', 'Elyscia', 0, 'M', '2010-05-03', 'American Indian or Alaska Native', 'Bachelor\'s degree', 'pr2ess@yourdomain.com', '16808851773');
INSERT INTO `patient` VALUES ('10020', 'Typhenete', 0, 'M', '2001-07-09', 'American Indian or Alaska Native', 'Associate degre', 'h12345i@example.com', '16864974651');
INSERT INTO `patient` VALUES ('10021', 'Gifford', 0, 'F', '2003-05-06', 'Hispanic or Latino', 'High school diploma or equivalent', 'seres@yourdomain.com', '11922096352');
SET FOREIGN_KEY_CHECKS = 1;

2. Insert the data into table Pancreatic_Cancer

INSERT INTO `pancreatic_cancer` VALUES ('10001', '0', 0, '2018-01-03');
INSERT INTO `pancreatic_cancer` VALUES ('10002', '0', 1, '2019-04-09');
INSERT INTO `pancreatic_cancer` VALUES ('10008', '1', 0, '2019-05-19');
INSERT INTO `pancreatic_cancer` VALUES ('10009', '0', 0, '2019-03-20');
INSERT INTO `pancreatic_cancer` VALUES ('10011', '1', 1, '2018-05-30');
INSERT INTO `pancreatic_cancer` VALUES ('10016', '1', 0, '2018-04-21');
INSERT INTO `pancreatic_cancer` VALUES ('10017', '0', 1, '2019-05-05');

3. Insert the data into table IA_Test_Record

INSERT INTO `ia_test_record` VALUES ('10001', 5, 0, -1, 1, '2018-01-03', 18.5, 111, 80, 1, 1);
INSERT INTO `ia_test_record` VALUES ('10002', 3, 1, 5, 1, '2019-04-09', 23, 119, 80, 0, 1);
INSERT INTO `ia_test_record` VALUES ('10003', 18, 1, 3, 1, '2018-04-05', 22.8, 120, 82, 0, 0);
INSERT INTO `ia_test_record` VALUES ('10004', 11, 1, 7, 1, '2018-04-05', 25.6, 145, 92, 1, 1);
INSERT INTO `ia_test_record` VALUES ('10005', 1, 1, 3, 1, '2019-06-10', 24.3, 180, 110, 0, 1);
INSERT INTO `ia_test_record` VALUES ('10006', 16, 0, 1, 1, '2020-01-30', 21.8, 165, 107, 1, 1);
INSERT INTO `ia_test_record` VALUES ('10007', 5, 0, 14, 1, '2020-06-15', 22.7, 125, 82, 0, 1);

Run and Test

1. New Patient Registration

Input the personal information into the Patient Information. Using the SQL code:
INSERT INTO `patient` VALUES ('10022', 'Iya', 0, 'F', '2010-03-01', 'American Indian or Alaska Native', 'Bachelor\'s degree', 'IYA@gmail.com', '16777851773');

2. Patient Information Query

Input the patient's first name, last name and birth date, we can get the personal information of the patient and also the patient id for the further query. Using the SQL code:
*SELECT * FROM patient*
WHERE first_name = "Malyn" AND last_name = "Ayleth" AND Birth_Date = "1997-5-6"

3. Patient Exam Query

Input the patient id into the Patient Information. Using the SQL code:
*SELECT * FROM IA_Test_Record*
WHERE Patient_id = 10010

Or we can input the patient's first name, last name and birth date, we can get the exam information of this patient. Using the SQL code:
*SELECT * FROM patient*
WHERE patient_id in (SELECT patient_id FROM patient WHERE first_name = "Malyn" AND last_name = "Ayleth" AND Birth_Date = "1997-5-6")

4. New Exam Update

Input the new exam information into the Patient Information. Using the SQL code:
INSERT INTO `ia_test_record` VALUES ('10001', 6, 0, 0, 1, '2018-02-03', 18.7, 110, 81, 1, 1);

Machine Learning Algorithm

With the Electronic Health Record (EHR) system, we can keep collecting the patients with these two diseases, Pancreatic Cancer and Inflammatory Arthritis, as a dynamic database which is expected to be a meaningful system.

The database is supposed to work in at least two ways. For the people who still haven't had these two diseases, this system is used to predict whether they are possible to get these two diseases. And for patients who have already had one of these two diseases, this system works as an exploratory database to detect the probability of having the other one disease.

For example, we first build a model with predictors BMI, SBP, DBP, smoking status, PainDETECT-score and response disease inflammatory arthritis. With this predictor model, once the information of new people for these 5 predictors, we are able to know whether these new patients would get inflammatory arthritis. Meanwhile, if the people above are pancreatic cancer patients, then we could also know if they are possible to have inflammatory arthritis.

For the model, we constructed a deliverable Python code template for further using which could be found in supplements. The predictive models include Logistic Regression model, Support Vector Machine (SVM), Random Forest, Ridge Regression model and K-Nearest Neighbor (KNN) model. In practice, all of these five models are expected to be built following by selecting the reasonable one. The parameters needed for the predict model are explained in the Python as comments.

Discussion

An Electronic Health Record (EHR) System and corresponding deployment Python templated code were performed in this project. With these initial steps, we are supposed to take advantage of the existing data for pancreatic cancer patients and inflammatory arthritis patients and know more about the relationship between these two diseases as well as getting an exploratory diagnostic model for both two conditions. Since this study aims to give out an initial framework for further development, for this project, we didn't provide the validation for the predictive model which might make the model less convincing.

This EHR system has the ability to keep updating new information of existing patients and adding new registration patients which obviously can include more diseases following by becoming a huge and useful database for predicting diseases as well as exploring the relationship with regard to more different diseases instead of only pancreatic cancer and inflammatory arthritis patients.

As more data is included, the application of this system is also supposed to be enriched. However, the validation of predictive models relates to not only the appropriation of the predictors but also the strength of the relationship between two diseases. This system should act as a platform of obtaining data for different purposes and researches and more consideration should be involved when querying data.

In a nutshell, based on the framework of this study, we are looking forward to developing a strong database for further research.

Reference

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30. doi:10.3322/caac.21442

[2] Padoan A, Plebani M, Basso D. Inflammation and Pancreatic Cancer: Focus on Metabolism, Cytokines, and Immunity. Int J Mol Sci. 2019;20(3):676. Published 2019 Feb 5. doi:10.3390/ijms20030676

[3] Virchow, R. Die Krankhaften Geschwülste; Springer: Berlin, Germany, 1863.

[4] Patel, R.; Lara, S.; Johnson, J.; Kulkarni, P. New pancreatic adenocarcinoma in a Crohn's patient treated
with tumor necrosis factor (TNF) inhibitors for 6 months. J. Gastrointest. Cancer 2014, 45, Suppl 1. 226–229.
[5] Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin. 2006 Oct;22(10):1911-20. doi: 10.1185/030079906X132488. PMID:17022849.
[6] Rifbjerg-Madsen S, Christensen AW, Christensen R, Hetland ML, Bliddal H, Kristensen LE, Danneskiold-Samsøe B, Amris K. Pain and pain mechanisms in patients with inflammatory arthritis: A Danish nationwide cross-sectional DANBIO registry survey. PLoS One. 2017 Jul 7;12(7):e0180014. doi: 10.1371/journal.pone.0180014. PMID: 28686639; PMCID: PMC5501437.