DISENTANGLED REPRESENTATIONS LEARNING FOR COVID-19 SEQUELAE PREDICTION

by

Zhaorui Liu

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Computer and Information Technology

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Baijian Yang, Chair

Department of Computer and Information Technology

Dr. Jin Wei-Kocsis

Department of Computer and Information Technology

Dr. Petros Drineas

Department of Computer Science

Approved by:

Dr. John A. Springer

ACKNOWLEDGMENTS

Throughout the writing of this thesis, I have received a great deal of support and assistance. I wish to gratefully acknowledge my thesis committee for their insightful comments and guidance and my family for their support and encouragement.

First, I would like to express my great gratitude to my advisor, Professor Baijian Yang, who guided me in looking for the research question, studying the problem, and developing my own solution to the problem. Your perceptive feedback helped me to choose the right direction and successfully complete my thesis.

In addition, I would like to thank Dr. Jing Su, Dr. QianQian Song, and Dr. Tongling Zhang, for their insightful guidance that helped me dive into the research question. Thank you for providing the dataset used in my thesis and guiding me to develop my approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE2	Human Angiotensin-Converting Enzyme 2
AI	Artificial Intelligence
AKI	Acute Kidney Injury
CKD	Chronic Kidney Disease
CT	Computed Tomography
DCG	Discounted Cumulative Gain
DL	Deep Learning
DVAE	Disentangled Variational Auto-Encoder
ML	Machine Learning
EMR	Electronic Medical Record
NDCG	Normalized Discounted Cumulative Gain
SARS	Severe Acute Respiratory Syndrome
XAI	Explainable Artificial Intelligence

ABSTRACT

Severe acute respiratory syndrome (SARS)-CoV-2 emerged in late 2019, then became an unprecedented public health crisis. Hundreds of millions of people have been affected. What is worse, many researchers have revealed that COVID-19 may have long-term effects on varieties of organs even after recovery. Consequently, there is a need for the study of its sequelae. The purpose of this project is to use machine learning algorithms to study the relationship between patients' EMR data and long-term sequelae, especially kidney diseases. Inspired by a recent learning disentangled representation for recommendation work, this project proposes a method that (i) predicts the development trend of the kidney disease; (ii) learn representations that uncover and disentangle factors related to kidney diseases. The major contribution is that this model has high interpretability which enables medical works to infer the development of patients' condition.

CHAPTER 1. INTRODUCTION

This chapter gives an introduction of the research question. It introduces the background knowledge by stating the significance of this research question. Additionally, this study is based on some assumptions that are presented in this chapter. Besides, limitations and delimitations are important to define the scope of the study.

1.1 Background

Coronavirus disease 2019 (COVID-19) emerged in approximately December 2019, then rapidly affected the entire world. As of May 2021, more than 100 million individuals have been infected and it has caused more than 3 million deaths worldwide. This outbreak has already resulted in tremendous societal disaster in the United States, which caused over 33 million cases and over 590 thousand deaths according to Johns Hopkins Coronavirus Center (*COVID-19 Map*, n.d.). Although this pandemic is still not over, emerging issues have yet to be resolved. A lot of research has revealed that COVID-19 can cause long-term effects even after recovery. Martinez-Rojas et al. (2020) presented that in spite of the fact that COVID-19 mostly affects the respiratory system, its multi-organ involvement causes severe damage to many different organs as well, including the heart, liver, blood vessels, bone marrow, kidneys, gut, and brain, which could be related to Human angiotensin-converting enzyme 2 (ACE2). ACE2 is an entry receptor and commonly seen in lung, heart, and kidneys cells (Alimadadi et al., 2020).

COVID-19 is a contagious disease that possibly brings serious complicating diseases, and long-term sequelae of COVID-19 are unknown. So far, few works study long-term sequelae of patients survived the acute phase. Janiri et al. (2020) proposed an important open question: ”Once recovered from COVID-19, what will happen to patients, and how does the virus affect their bodies?” (Janiri et al., 2020). 22% of patients had acute kidney injury (AKI), but some studies have found that acute kidney injury was not caused by SARS-CoV2 infection itself (Wang et al., 2020). This pandemic has already resulted in tremendous societal disasters worldwide. A rough estimate is that more than 22 million people will suffer from AKI. Therefore, it is a huge challenge for the health care system.

Within the intense stage of COVID-19, the incidence of complications including liver and kidney is high, but it is not clear how many of these complications will persist in the medium and long term (Dawei Wang, 2020). It may be expected that patients with severe symptoms will have long-term complications and permanent damage. There is a lack of guidelines for postoperative recovery from COVID-19. "It is predicted that 45% of patients discharged from hospital will require support from healthcare and social care and 4% will require rehabilitation in a bedded setting" (Barker-Davies et al., 2020). Therefore, it is obviously necessary to propose insights for the recovery of COVID-19 survivors. there are very few guidelines on how to better rehabilitate such patients.

COVID-19 affects different individuals differently, from insignificant symptoms to the serious respiratory system infection requiring put on ventilator. Therefore, it is necessary to conduct further research on the sequelae of COVID-19. This project is based on the hypothesis that some patients infected with COVID-19 may still have potential long-term sequelae, especially kidney diseases, even after recovery. Furthermore, leveraging machine learning to analyze patients' clinical data can infer how they are different from other people, and on this basis make long-term treatment recommendations for patients.

The problem as perceived by the researcher at this point is that although other researchers have proposed some machine learning models that have high accuracy on related problems, however, interpretability of their models is the major concern when applied in real-world practice. Since the machine learning models are supposed to assist medical workers in making decisions about the patient's condition. If medical workers do not know why the machine learning model makes such predictions, they cannot trust these predictions. In the medical field, misdiagnosis is not acceptable. Imagine that if a patient suffers from a certain disease and the machine learning model predicts negative, this may cause them to miss the optimal treatment time; conversely, if a patient does not have some disease and the machine learning model gives a positive prediction, this can be a disaster. In this case, the practicality of the machine learning model is greatly reduced.

1.2 Problem Statement

As researchers and medical workers better understand COVID-19, the proportion of patients surviving the virus is increasing at present. However, having defeated the virus is just the beginning of an uncharted recovery path for infected patients. The long-term sequelae of COVID-19 still remain unknown. "The sequelae in those who survive this illness will potentially dominate medical practice for years and rehabilitation medicine should be at the forefront of guiding care for the affected population" (Barker-Davies et al., 2020). According to the European Centre for Disease Prevention and Control (ECDC), a correlation was observed between chronic kidney disease (CKD) and COVID-19 severity: In spite of the fact that Acute kidney injury (AKI) is not very commonly seen in the mild COVID-19 infection, it is more common in critically ill patients. "It is probable that kidney lesions acquired during the disease's activity remain as sequelae that may result in a slow and asymptomatic progression towards advanced stages and CKD" (Herrera-Valdés et al., 2020), which indicates that severely ill patients may still have medium- and long-term renal sequelae even after recovery.

The purpose of this project is to provide insights into the long-term sequelae of the kidneys of COVID-19 patients. This project aims to give recommendations to health care workers about potential sequelae that need careful attention (e.g., plan what is necessary in the medium and long term) in order to optimize future healthcare delivery. The final aim is to let the affected patients return to their normal life in a healthy state.

Since interpretability is a very significant component in this study, learning disentangled representations is an appropriate way to find the correlation between examination results and the development of kidney diseases. In this study, a Disentangled Variational Auto-Encoder (DVAE) will be built to learn a set of prototypes that illustrate the correlation between various examination results and the development of kidney diseases.

1.3 Research Question

This research contributes answers for following questions:

1. How to predict the Chronic kidney disease (CKD) progression risk base on clinical variables from EMR data?

2. Which of the clinical variables may have hidden relationship in the latent space?

3. How do clinical variables affect the the CKD progression?

1.4 Significance

The White House called global AI researchers to take action to develop novel data mining techniques to assist research related to COVID-19. Alimadadi et al. (2020) showed that a great number of machine learning techniques have been used to analyze the biochemical and clinical data shortly after the pandemic. Digging out hidden information from clinical data to improve treatments is promising. AI techniques are able to unearth implicit patterns and provide insights for health care workers. The development of advanced ML-based models to reveal unique mechanical insights which cannot be obtained from traditional methods, is essential for the future health care system.

Therefore, in summary, applying machine learning technology to study long-term sequelae of COVID-19 is of great significance and may inspire long-term treatments. However, this issue has not been studied and discussed so far. In consequence, an in-depth study will lead to an extreme meaningful discovery.

1.5 Assumptions

The assumptions for this study include:

- The data is true and reliable. Assume the dataset will be used in experiments comes from real electronic medical records (EMR) and assume the EMR data covers all the features needed in experiments.

- These indicators have some relationship with certain sequelae that cannot be explicitly detected, but machine learning models are able to catch the hidden patterns.

- It is possible to learn disentangled representations for latent factors that related to the sequelae.

1.6 Limitations

The limitations for this study include:

- There are a lot of missing data since a patient is very unlikely to do all the tests. In most cases, a patient may only do a few types of examinations.
- The performance of the proposed model is limited by the size of training data due to the limited computation power.

1.7 Delimitations

The delimitations for this study include:

- This project only studies the long-term sequelae of the kidney.
- This project aims to study various examination results using machine learning algorithms. The dataset will be used is from National COVID Cohort Collaborative (N3C). There is no interaction with patients.
- The purpose of this project is to conduct further research on the sequelae of COVID-19, provide insights for long-term treatments. However, this study will not produce any specific post-recovery treatments.

1.8 Summary

This chapter introduced the required background knowledge and defined the scope of the research question. Additionally, it presented the assumptions that the study was based on. It described the limitations and delimitations for the study as well. The next chapter gives a literature review of related works.

CHAPTER 2. REVIEW OF LITERATURE

The content of the literature review covers the study of the sequelae of covid-19, examples of utilizing the machine learning in the medical field, and the study of disentangled representation.

2.1 COVID-19 Sequelae

Some survivors have early sequelae related to COVID-19, e.g. cardiac sequelae (Demertzis et al., 2020). Some previous works have studied the clinical variables related to of COVID-19. In addition, the methods used to diagnose and treat patients have been widely studied as well. However, further studies on the potential sequelae in COVID-19 survivors remain to be conducted (Guan et al., 2020). It has been shown that patients might have phisical and mental diseases even though they survived severe acute respiratory syndrome (SARS) (Lam, 2009). Rogers et al. (2020) have studied the clinical variables of SARS and Middle East respiratory syndrome and presented that patients recovered from COVID-19 might have psychiatric sequelae. Nevertheless, there is a lack of investigations diving into the long term sequelae of patients who recovered from the intense stage of COVID-19. Besides, further studies on the treatment are also remain to be conducted (Janiri et al., 2020).

Within the intense stage of COVID-19, the incidence of complications including liver and kidney is high, but it is not clear how many of these complications will persist in the medium and long term (Wang et al., 2020). It may be expected that patients with severe symptoms will have long-term complications and permanent damage. There is a lack of guidelines for postoperative recovery from COVID-19. “It is predicted that 45% of patients discharged from hospital will require support from healthcare and social care and 4% will require rehabilitation in a bedded setting” (Barker-Davies et al., 2020).

Since there are many survivors of COVID-19, the sequelae may be of great significance for the survivors to get back to the normal life. It has been proved that the central nervous system (CNS) is a target affected by the coronavirus. Thus, it compromises the CNS and causes psychiatric diseases. A few affected patients showed intense psychiatric side effects such as delirium, encephalopathy and anosmia (Vaira, Salzano, Deiana, & De Riu, 2020). Lai, Ko, Lee,

Jean, and Hsueh (2020) showed that patients affected by COVID-19 tend to be more possible to have psychiatric diseases and when they get psychiatric diseases, the situation tends to be severer. Moreover, recent research works also mentioned that when patients suffer from mental diseases, the coronavirus could make their situation even worse. Nevertheless, there is no researchers have studied the correlation between mental diseases and the coronavirus.

In summary, the sequelae caused by COVID-19 infection is a real problem and remain to be resolved. Consequently, the problem studied in this project is of great significance.

2.2 Interpretability

People cannot trust anything if they don't understand why it works, especially in the medical field. Miller (2019) defined interpretability as: "Interpretability is the degree to which a human can understand the cause of a decision".

There is no doubt that machine learning models can learn patterns hidden in big data and make predictions with high accuracy. Nevertheless, most of the existing models do not have interpretability, which means people cannot explain the reason why they make such predictions. In this case, researchers are working on building ML models that people can understand assumptions and decisions behind predictions, that is, models are supposed to be open, transparent, and understandable.

2.2.1 Prediction-level Interpretation

Prediction-level approaches is used to understand the importance of each feature for each prediction. In order to achieve the prediction-level interpretation, a widely used method is to a relevance factor to each feature. In this case, a variable that has a higher relevance score represents a higher contribution to a certain prediction result (Murdoch, Singh, Kumbier, Abbasi-Asl, & Yu, 2019). For instance, a heat map can be used to represent the importance of each pixel contributing to the prediction.

2.2.2 Model-based Interpretation

"One type of post hoc model-specific explanation methods is knowledge distillation, which is about extracting knowledge from a complex model to a simpler model (which can be from a completely different class of models)" (Carvalho, Pereira, & Cardoso, 2019). There are a number of methods to achieve knowledge distillation, e.g. model compression (Polino, Pascanu, & Alistarh, 2018), tree regularization (Wu et al., 2018), etc.

2.2.3 Model-agnostic Interpretation

The contribution of input to output can be determined through occlusion and omission. This line of work "tries to answer the question: which parts of the input, if they were not seen by the model, would most change its prediction? Thus, the results may be called counterfactual explanations" (Du, Liu, & Hu, 2020). However, it is not practical in the real world since few models take blank inputs. Another problem is that it may cause side effects. For example, if a number of pixels are occluded with green color, it may give a prediction of a lawn class which is obviously a disturbance and even cause the misclassification.

2.2.4 Applications

An important application area of Interpretability is model debugging. A typical example is Adversarial Machine Learning (Nguyen, Yosinski, & Clune, 2015). As Du et al. (2020) stated, ML Models might give wrong predictions for specially crafted adversarial examples. However, these errors can be easily discovered by humans. Interpretability helps humans analyze why models fail and furthermore promote robustness.

Another significant application is knowledge discovery. Humans are able to have a better understanding about machine learning models make a certain decision. In this case, the users with expertise in the applied area may be able to help justify the prediction. It is even possible to obtain new insights that were not noticed in the past. Thus, new theories are derived from there (Du et al., 2020). For instance, Caruana et al. (2015) proposed an interpretable model to predict

the mortality risk for pneumonia patients. Health care workers provide more aggressive treatment based on the insights provided from the prediction. Facts have proved that the treatment achieved better results.

In summary, interpretability enables people to have a deeper understanding of the model. Humans can explain why it happens, which means the thinking procedure has meanings to human beings. Even if the models don't give the right predictions, there is still some useful information. As a result, a model with interpretability is much more valuable than that only giving predictions.

2.3 Machine Learning in the Medical Field

Machine learning techniques have been widely studied and applied to many fields since they are able to reveal the hidden information embedded in the raw data. Many researchers have studied applying machine learning into the medical field:

Xiao et al. (2020) proposed a deep learning model to estimate the disease severity of COVID-19 patients based on computed tomography (CT) imaging. It used multiple instance learning. They used the data of 303 patients in the People's Hospital of Honghu to train their model and tested it on the data of 105 patients in The First Affiliated Hospital of Nanchang University. They evaluated their model by calculating the receiver operating characteristic curve and the confusion matrix. On the training set, the accuracy was over 97% and the area under the curve (AUC) was 0.987. While on the test set, the accuracy was over 81% and the AUC was 0.892. The analysis on the subgroup of patients without server symptoms on admission, accuracies in the Honghu and Nanchang subgroups were 97.0% and 81.6% and the model achieved ACUs of 0.955 and 0.923 respectively.

Booth, Abels, and McCaffrey (2020) conducted a review assessing research facility information and mortality from patients with positive RT-PCR test for SARS-CoV-2. They created a machine learning model utilizing 5 serum chemistry laboratory parameters from 398 patients for predicting patient expiration status. Their model achieved over 90% on both sensitivity specificity for the prediction of death on test data.

Patel et al. (2021) developed a ML model based on radionics that analyses CT images and clinical variables to predict COVID-19 severity as well as the possible of deterioration to severe

diseases in the future. They collected data from patients confirmed to be positive for COVID-19. Moreover, they collected radiomics features from patients' chest CT images. They trained 2 models: one for prediction severity and the other for predicting the progression to severe diseases. These models were trained with radionics features and clinical variables. They evaluated their models by calculating the receiver operating characteristic area under the curve (ROC-AUC), concordance index (C-index), and time-dependent ROC-AUC and compared the results with consensus CT severity scores made by radiologists using visual evaluation.

Berenguer et al. (2020) proposed an interpretable machine learning model based on a semi-supervised classifier that utilizes a VAE to extract embeddings. They have optimized the 2 networks that take CT images as inputs. First, they developed a new conditional variational autoencoder (CVAE) that has a particular architecture with which the class labels can be integrated into the encoder. Second, they implemented a supervised CNN classifier taking advantage of the encoder structure.

In summary, applying machine learning in the medical field helps health care workers diagnose diseases. However, these models cannot be fully trusted because they are not interpretable. It is the research gap this project will fill up.

2.4 Variational Auto-Encoder

An auto-encoder contains two connected networks: one is encoder, the other is decoder. The encoder network receives the input and compresses it into a relatively small dense vector which can be converted back to the original input by the decoder. The encoder and the decoder are commonly trained together. The loss function is usually used to measure the difference between the reconstruction and the original data, called reconstruction loss. The encoder needs to select the most useful information in order to achieve a small size of the representation, while the decoder needs to learn how to extract as much as possible information from the representation and reconstruct the original input with it. Together they form an auto-encoder.

Standard auto-encoders learn to compress data and reconstruct from the encoded data, but apart from being used in some applications, such as denoising auto-encoders, their usage is quite limited. The basic problem with auto-encoders is that they convert their input into a code vector,

the latent space in which they are located may be discontinuous, or allow simple interpolation. A generative model is not build to reconstruct the output exact same as the input. It is expected to generate variations that are derived from the original input by altering some dimensions of factors in the latent space.

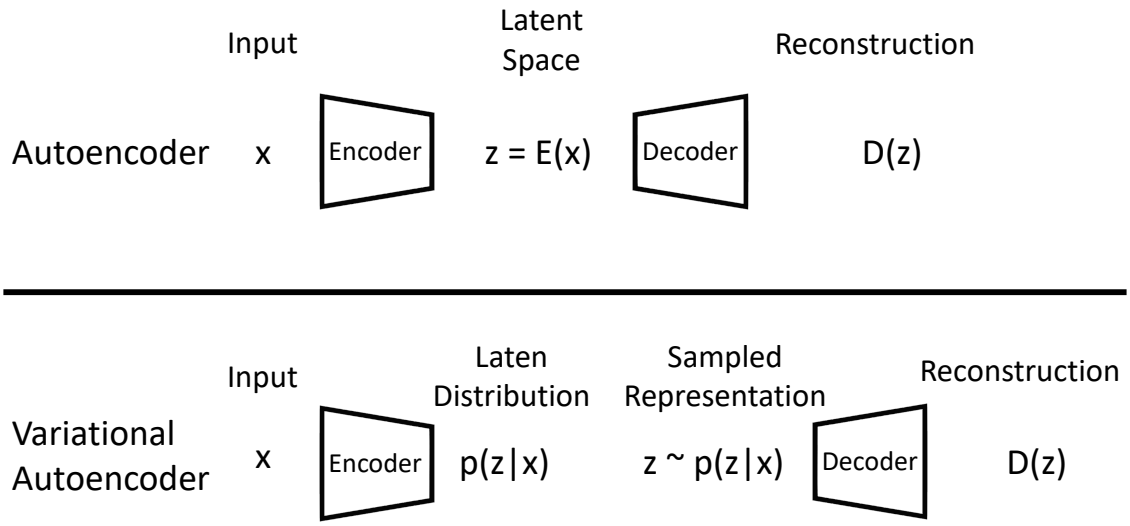

Figure 2.1. Difference between auto-encoder (deterministic) and variational auto-encoder (probabilistic)

Variational auto-encoders (VAEs) have a unique property: their latent space is continuous, which enables random sampling and interpolation (Doersch, 2016). It is achieved by making some constraints: the encoder does not output a code vector of size n, but outputs two vectors of size n: the average vector μ and another standard deviation vector σ. The difference between auto-encoder and variational auto-encoder is shown in Figure 2.1.

The training process is as follows (Kingma & Welling, 2014):

1. The encoder network takes the input and encode it as a distribution over the latent space.
2. Sample a point from the distribution calculated in the previous step in the latent space.
3. The decoder network takes the point as input and reconstruct the original input.
4. The reconstruction error is computed and backpropagated through the network.

2.5 Learning disentangled Representations

Each dimension in the disentangled representation represents a dimension of factors that contain meaning information to human beings (Gilpin et al., 2018). Currently, there is no formal definition of disentangled representation, but the generally accepted informal definition is as follows:

- A disentangled representation ought to isolated the particular, informative components of varieties in the data.
- A change in a certain factor leads to a variation in latent space while any other changes of different factors contribute very little (Bengio, Courville, & Vincent, 2013).
- An alter in an embedding factor of the variable z_i ought to lead to an alter in a factor in the encoded representation $r(x)$ (Locatello et al., 2019).

Intuitively, the benefits of disentangled representation are:

- They ought to display the data in a compact format which has the interpretability (Chen et al., 2016).
- They ought to be valuable for transfer learning (Chen et al., 2016).

To sum up, the purpose of disentanglement is to separate latent, underlying, high-level explanatory factors from low-level observation data. The purpose of disentangled representation is to spparate factors independently and map them to laten units of different dimensions in representation vectors. On the one hand, such a representation can integrate a variety of explanatory factors, making it more semantic and explanatory; on the other hand, the independence between different factors is robust, and a slight disturbance of a factor will not affect other factors. If a task is highly related to a certain factor , the separated representation can better solve the task through the related factor, which is conducive to transfer learning and small sample learning. On the contrary, since entangled representation couples all the factors, the learned representations are highly dependent on the specific task, and the interpretability is not strong enough.

2.5.1 Learning Disentangled Representations for Recommendation

Ma, Zhou, Cui, Yang, and Zhu (2019) developed an advanced VAE named MACRo-mIcro Disentangled Variational Auto-encoder (MacridVAE). This model was used to learn disentangled representations and give recommendations based on that. Their approach "achieves macro disentanglement by inferring the high-level concepts associated with user intentions (e.g., to buy a shirt or a cellphone), while capturing the preference of a user regarding the different concepts separately. A micro-disentanglement regularizer, stemming from an information-theoretic interpretation of VAEs, then forces each dimension of the representations to independently reflect an isolated low-level factor (e.g., the size or the color of a shirt)" (Ma et al., 2019).

The main objectives of this work are (1) to learn a factorized representation of user's interests about item classes; (2) to disentangle users' favourite at a low level, e.g. colors, sizes, etc. The intuition of the two objectives is that (1) users may have varied interests, e.g., clothes, electronic devices; (2) users' preference about each product may depend on the categories itself, that is, a user who prefers a red handbag may not like a red refrigerator (Ma et al., 2019).

The decoder predicts the rank of items are likely to be interacted with a user while the encoder computes the representation of a user given behavior data.

Their experiments are conducted on five real-world datasets, four of which are the large-scale Netflix Prize dataset (Bennett, Lanning, et al., 2007), and 3 different sizes MovieLens datasets, i.e., ML-100k, ML-1M, and ML-20M (Harper & Konstan, 2015). The other dataset is AliShop-7C which is from Alibaba's online shopping platform. It consists of associations from users to merchandises. There are seven categories of products in the dataset. Each item has attributes such as titles and images. Each user has at least two interactions associated with items from different categories.

There are two state-of-the-art mothods used in their experiments, i.e., MultDAE (Liang, Krishnan, Hoffman, & Jebara, 2018) and β-MultVAE (Liang et al., 2018). Notice that β-MultVAE does not learn disentangled representations.

They follow the experiment design established by Liang et al. (2018). They evaluate the performance of all three models under strong generalization (Marlin, 2004):

1. Split all users into training set, validation set, and test set.

2. Train models using all the interactions of the training set.

3. To evaluate, split the validation set or the test set into 2 parts: a training set and a test set, i.e. validation-training set and validation-test set or test-training set and test-test set. Learn users' preference from the training set and then evaluate how well models perform on the rest of the unseen dataset, i.e. the test set.

They use the same evaluation metrics as Liang et al. (2018) which are Recall@R and NDCG@R.

2.6 Summary

This chapter reviewed a number of the literature relevant to the sequelae of COVID-19, machine learning techniques applied in the medical field, variational auto-endoders and disentangled representation learning. The next chapter describes the methodology to be used in this project.

CHAPTER 3. PROPOSED SOLUTION

This chapter provides the proposed framework. The dataset used in this study, the data preprocessing procedure, the evaluation criteria, and the measure to handle the imbalanced data are introduced in this chapter.

3.1 System Modeling

This section describe the framework of the system. It traduces the necessity of developing interpretable ML models in the medical field. In addition, it shows the raw dataset, the data preprocessing procedure and the structure of the model.

3.1.1 Introduction

Predicting the development of the disease based taking advantage of machine learning models is very promising, and researchers have proposed many models with very impressive accuracy. Since machine learning is a black box, that is, although the model can make predictions, people don't know why it makes such predictions. In the medical field, these predictions are related to the health and even life of patients, so the interpretability of the model is of great significance. Therefore, the project aims to study a machine learning model with high interpretability. On the one hand, it predicts the development of the disease based on the patient's EMR data, and on the other hand, disentangles the latent factors.

3.1.2 Research Approach

The research approach is mainly divided into three parts: the first step is data preprocessing, in this step raw data is converted into the input format accepted by the machine learning model; the second part is building the model, in this step the model and the loss function are designed to learn the latent factors of the data; the third step is to evaluate the model and interpret the pattern learned by the model.

3.1.2.1 Data Preprocessing

Raw data is based on encounters, that is, every record in the database is an encounter record. Each encounter record contains the patient's, gender, feature name, corresponding data, and time. The features used in this project are shown in Table 3.1. The dataset used in this project includes patients of different ages, races, sexes, and health conditions. A total of 20 features are used for research. In addition, the patient's survival time can be obtained according to the outcome database. First, according to the predetermined baseline time, obtain the data two years before the baseline time, calculate the median value of each patient's clinical variables, and also calculate the BMI index based on the height and weight. Second, according to the classification standards of various clinical variables commonly used in medical fields, each indicator is mapped to the corresponding level. In addition, due to the missing data also has some certain meaning (for example, a patient who does not have high blood pressure may have very little or even on blood pressure data, however a patient with high blood pressure is likely to have a great number of blood pressure testing encounters), it is necessary to add a category to represent the absence.

Table 3.2 gives an example of a record in the ERM dataset. The column "concept.id" contains the name of the feature and the column "nval.num" contains the corresponding value. Note that the age is calculated based on the date of the exam and their birthday.

There is another dataset called "outcome" indicating how soon the condition becomes worse. Table 3.3 gives an example.

As described above, calculate means, classify them, and combine the EMR dataset with the outcome data set to get the final table shown in Table 3.4. The value $x_{i,j}$ in the ith row and jth column represents that the ith patient has jth state. A state can be a certain class of a clinical variable, e.g. EGFR_0. Note that 0 means the patient does not have any records of this clinical variable.

The final table is an adjacent matrix in fact, hence a graph can be built from it. Each node in the graph represents a patient or a state, and an edge connecting a patient and a state represents that the patient has this state. Since most of the clinical variables are null, the graph is a sparse graph as illustrated in Figure 3.1.

The dataset is then divided into three sub-datasets (e.g., training set, validation set and testing set).

Table 3.1. *Features in EMR data*

Feature	
ALK	Alkaline Phosphatase
ALT_SGPT	Alanine Aminotransferase
AST_SGOT	Aspartate Aminotransferase
BP_DIASTOLIC	BP_DIASTOLIC
BP_SYSTOLIC	BP_SYSTOLIC
CHOLESTEROL	CHOLESTEROL
CREATINE_KINASE	CREATINE_KINASE
EGFR	Creatinine
HBA1C	Hemoglobin A1c
HDL	High Density Lipoprotein
HEMOGLOBIN	HEMOGLOBIN
HT	HEIGHT
INR	Prothrombin Time and International Normalized Ratio
LDL	Low-density Lipoprotein Cholesterol
TBIL	Total bilirubin
TRIGLYCERIDES	TRIGLYCERIDES
TROPONIN	TROPONIN
WT	Weight
Age	

Table 3.2. *EMR dataset example*

pat_id	concept.cd	nval.num	units.cd	age	csn
.....	.....	.....	.....	.....	.....
64146	EGFR	66.0821139111329	$mg/mL/1.73m^2$	84.6657534246575	30857975
.....	.....	.....	.....	.....	.....

3.1.2.2 Learning Disentangled Representations

After the dataset is preprocessed, the next step is to build a Disentangled Variational Auto-Encoder (DVAE) that learns the hidden patterns embedded in patients' EMR data. The DVAE is expected to learn disentangled representations and they should have meaning to human beings. In addition, the disentangled representations are also supposed to be controllable by manipulating the encoded vectors. DVAE is mainly composed of two components, one is an

Table 3.3. *Outcome dataset example*

pat_num	year2_ture	year2_false	year5_true	year5_false	year10_true	year10_false
.....	.....	.....	.....	.....	.....	.....
63655	1	0	1	0	1	0
.....	.....	.....	.....	.....	.....	.....

Table 3.4. *Final table example*

pat_num	.....	$EGFR_1$	$EGFR_2$	.....	ALK_0	ALK_1	.....	year2_true	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
63655	.....	0	1	.....	1	0	.....	1	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

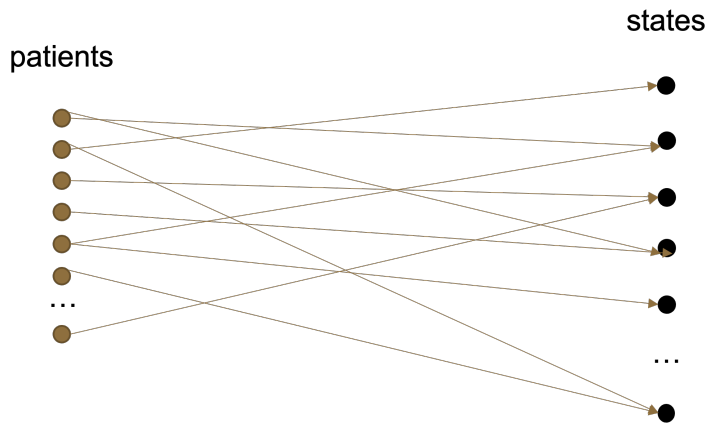

Figure 3.1. Building a sparse graph from the final table

encoder and the other is the decoder. The input data is the graph constructed in the previous step. The graph is mapped to the latent space through the encoder network, and the decoder reconstructs the graph with the value of the latent space. Finally, calculate the reconstruction loss based on the original graph and the reconstructed graph. Then adjust the parameters in the network according to the output of the loss function. A set of prototypes are learned from the relationship of patients' states and outcomes. The are expected to capture hidden patterns from patients' state to final outcomes.

The next step is to use the training set to train a DVAE and use the validation set to prevent overfitting during the training process. The training process is shown in Figure 3.2. The input is the sparse graph built from the final table. The DVAE model first encodes the input into the latent space and decodes the representation to reconstruct the sparse graph, then evaluates the construction loss based on the input and the output. Finally, use the testing set to evaluate DVAE performance.

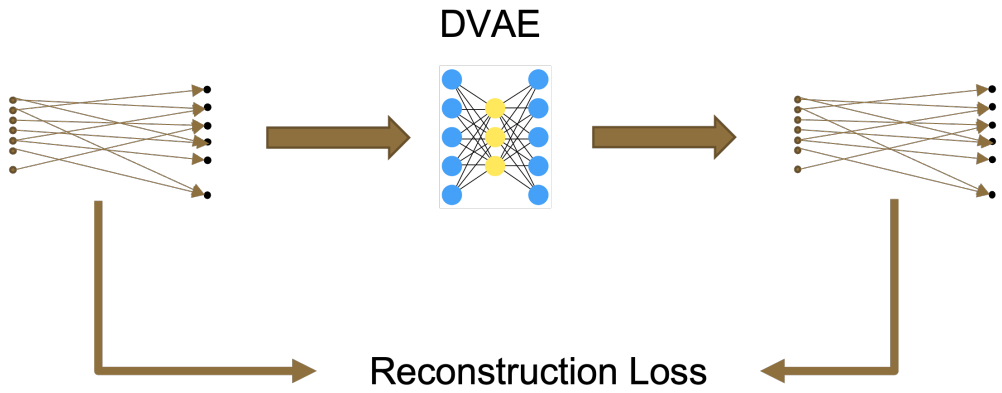

Figure 3.2. Training the DVAE model

Note that the originally DVAE gives out a relevance score associated with each item and relevance scores lies in the range $(-\infty, 0)$. As a consequence, there is an extra step to normalize relevance scores. A widely used normalization function: $r(i) = \frac{r(i)-min}{max-min}$ was used in this step, where $r(i)$ describes the relevance of item i, max is the maximum relevance score and min is the minimum relevance score.

3.2 Evaluation Criteria

As mentioned in Section 2.5.1, Liang et al. (2018) used NDCG@R as evaluation criteria in their experiments and NDCG (Normalized Discounted Cumulative Gain) is commonly used to measure how good a recommendation system is. It mainly measures two aspects:

- The degree of correlation between the series of recommended items and the users' real interests, which means how would a user like them given a set of recommended items.

- How well does the order of the recommended items match users' real interests? The reason is obvious, most users would look through recommended items from the top to the bottom.

However, the order does not matter in this project. As a result, different from the evaluation metrics used in Liang et al. (2018), the researcher used cross entropy loss and accuracy as evaluation metrics for the following reasons:

- NDCG@R and Recall@R are not suitable under this circumstance as they not only consider the relevance of the prediction and the ground-truth, buy also take the order of predicted items into consideration as well. However, the order does not matter under such situation.
- Cross entropy loss is normally used to measure how two probabilities differ from each other. After the normalization step mentioned in Section 3.1.2.2, relevance scores lie in the range $[0, 1]$ which can be considered as probabilities.

3.3 Handling the Imbalanced Data

Due to the reason that the outcome dataset is highly imbalanced, the model would tend to predict one class over the others. Table 3.5 shows the summary of the validation set and the test set. The sizes of validation set and test set are both 458. It is easy to see 2 characters:

1. The number of non-empty records in 2 years is much greater than the number of non-empty records in 5 years, and the number of non-empty records in 5 years is much greater than the number of non-empty records in 10 years.
2. The number of false is much greater than the number true for all 3 classes.

The above characters would cause the class $year2\_false$ dominate the prediction. In order to solve this problem, weights need to be added to each class. The weight of each class can be calculated by:

$$w(i) = \frac{1}{n_i} \times \frac{N}{M} \tag{3.1}$$

Table 3.5. *Summary of the validation set and the test set*

	year2_ture	year2_false	year5_true	year5_false	year10_true	year10_false
Validation set	34	424	54	262	80	97
Test set	60	398	105	219	136	78

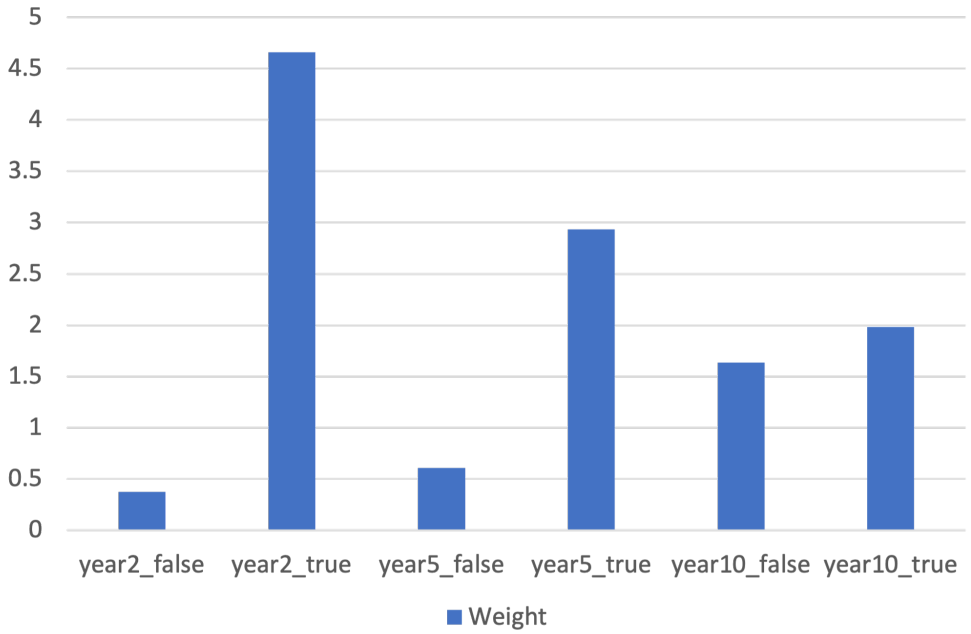

Figure 3.3. Weights of each class

where $w(i)$ is the weight of the ith class. n_i is the number of occurrence of the ith class. N is the sum of n_is and M is the number of classes. Figure 3.3 illustrates weights of each class to balance the impact.

Additionally, downsampling is an alternative method that randomly picks a similar size of training data from the larger class (Provost, 2000). In this way, the dataset becomes balanced and no class dominates the prediction.

3.4 Summary

The proposed methodology and evaluation criteria are described in this chapter. Due to the imbalanced data, techniques to prevent one class dominating the prediction is also introduced. The next chapter shows empirical results.

CHAPTER 4. EXPERIMENTS

This chapter provides experiments design and empirical results. The experimental setup, including the environment, the baseline model and evaluation metrics are introduced in this chapter. In addition, empirical results are given and the interpretability is disscussed.

4.1 Experimental Setup

This section describe the experimental setup, including the environment, hyper-parameters, the baseline model and evaluation metrics used in experiments.

4.1.1 Environment

All the experiments were run on a personal laptop, 2 2017 Apple MacBook Pro with a 2.3 GHz Dual-Core Intel Core i5 processor, 8GB memory and 256GB SSD running on MacOS 12.0. The DAVE model was built based on Tensorflow 1.14.0. The DVAE was trained using CPU-only. The Python version was 2.7.6, running in an Anaconda virtual environment.

4.1.2 Hyper-parameters

Table 4.1 shows the hyper-parameter values that gave the best results in experiments. Since some of them affected the accuracy and the cross entropy loss significantly, the researcher chose a set of values for some of them and evaluated the different combinations.

Considering the fact that all experiments were run on a personal laptop using CPU only, it took less than 10 minutes to train all the models. The time efficiency of the model was acceptable.

Table 4.1. *Values of hyper-parameters*

Parameter	Meaning	Value
epoch	Number of training epochs	200
batch	Training batch size	500
lr	Initial learning rate	10^{-3}
rg	L2 regularization	0
keep	Keep probability for dropout	0.5
beta	Strength of disentanglement	0.2
tau	Temperature of sigmoid/softmax	0.1
std	Standard deviation of the Gaussian prior	0.075
kfac	Number of facets (macro concepts)	7
dfac	Dimension of each facet	100
nogb	Disable Gumbel-Softmax sampling	False
seed	Random seed	98765

4.1.3 Baseline

The DVAE was compared with a random forest classifier. In order to fit a random forest classifier, the dataset needs some transformation. Similar to 3.1.2.1, the raw data needs to be processed in the following steps:

1. For each type of examination of each patient, filter all the data points after the baseline age and calculate the mean.

2. According to classification standards, classify each mean value calculated in the previous step to a certain class.

3. Fill out any empty class with 0.

4. Classify the outcome vectors.

The hyper parameters used in the random forest model were set to default except the max_depth was set to 30. The default values are given in Table 4.2.

Table 4.2. *Default parameter values used in the random forest*

Parameter	Default value
n_estimators	100
criterion	gini
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0.0
max_features	auto
max_leaf_nodes	None
min_impurity_decrease	0.0
bootstrap	True
oob_score	False
n_jobs	None
random_state	None
verbose	0
warm_start	False
class_weight	None
ccp_alpha	0.0
max_samples	None

4.1.4 Evaluation Metrics

4.1.4.1 Accuracy

One important criteria is accuracy. Accuracy of each class, i.e. year2, year5, year10, and overall accuracy were measured. The method used for calculating the accuracy was defined as:

Considering year2, year5 and year10 are 3 separate classes and for each:

$$acc = \frac{Number\ of\ right\ predictions}{Number\ of\ patients} \tag{4.1}$$

4.1.4.2 Cross Entropy Loss

As mentioned in Section 3.2, cross entropy loss was used to measure the difference between predictions and groundtruths. For the prediction vector p and the groundtruth vector q, the cross entropy loss was calculated by:

$$L(p,q) = -\sum_{x} p(x) \log q(x) \tag{4.2}$$

Confusion Matrix, precision, recall, and F-1 score were evaluated in experiments as well.

4.2 Results

4.2.1 Empirical Results

Figure 4.1 illustrates training logs. The cross entropy loss was less than 0.15 on the validation set and droped after around 100 epochs. On the test set, the accuracy of each class and the overall accuracy are given in Table 4.3.

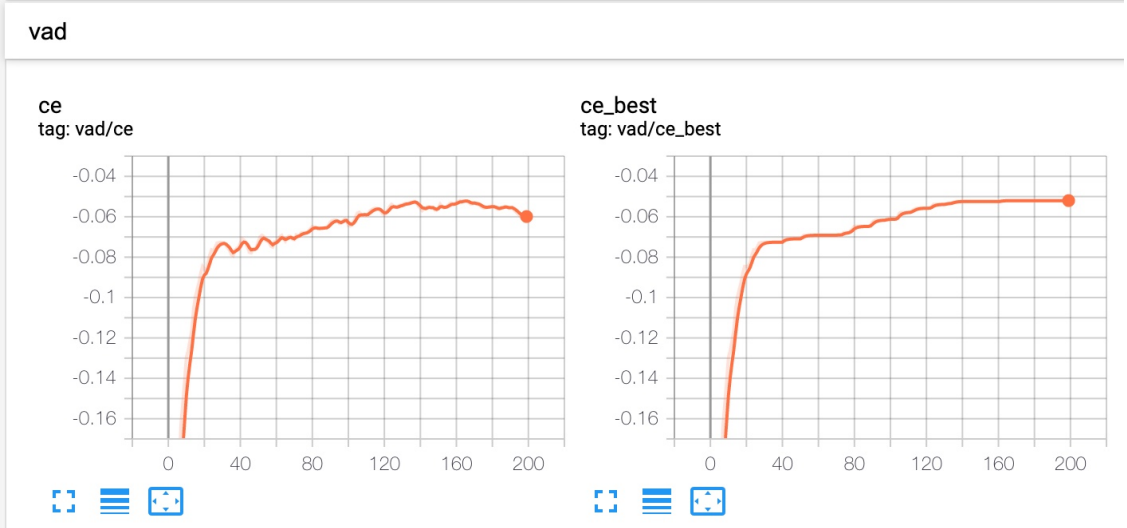

Figure 4.1. Tensorboard training logs

As mentioned in Table 3.5, year2 data points are much more than year5 data points and year5 data points are much more than year10 data points. It is reasonable that the accuracy of year2 class is the highest. Besides, because of the same reason, even though the accuracy of year10 class is pretty low, the overall accuracy is still acceptable.

As for applying downsampling to balance the dataset, due to the lack of training data, the accuracy was not desirable. Since adding weights outperformed downsampling, the following experiments were based on the adding weights approach.

Table 4.3. *The accuracy of adding weights and downsampling*

Class	Accuracy (adding weights)	Accuracy (downsampling)
Year2	0.86900	0.59639
Year5	0.73148	0.48795
Year10	0.86916	0.38554
Overall	0.87363	0.48996

As mentioned previously, hyper parameters used in tuning the model affected the accuracy significantly. A set of values were selected to find the model with the best performance. Table 4.4 shows values of parameters and corresponding results. When one parameter varied, all other parameters were fixed.

Table 4.4. *The accuracy of using different hyper parameters*

	Epoch			Learning rate		
	50	100	200	$1e^{-4}$	$1e^{-3}$	$1e^{-2}$
Year2_acc	0.84061	0.86900	0.86900	0.79258	0.86900	0.85153
Year5_acc	0.66975	0.67593	0.73148	0.63889	0.67593	0.63580
Year10_acc	0.38785	0.57944	0.86916	0.37850	0.57944	0.72897
Overall accuracy	0.77111	0.81441	0.87363	0.72999	0.81441	0.81732
Cross entropy loss	0.15886	0.12753	0.08999	0.22127	0.12753	0.09038
Training time (seconds)	142.03	302.16	524.94	313.38	302.16	203.732

Figure 4.2, Figure 4.3 and Figure 4.4 visualize the prediction results on test set. Each row represents a patient. Rows are hierarchical clustered. In Figure 4.2, it is noticeable that for the vast majority of patients, the probability of being predicted to be false is greater than the probability of being predicted to be true. Besides, as time increases, the confidence level also decreases.

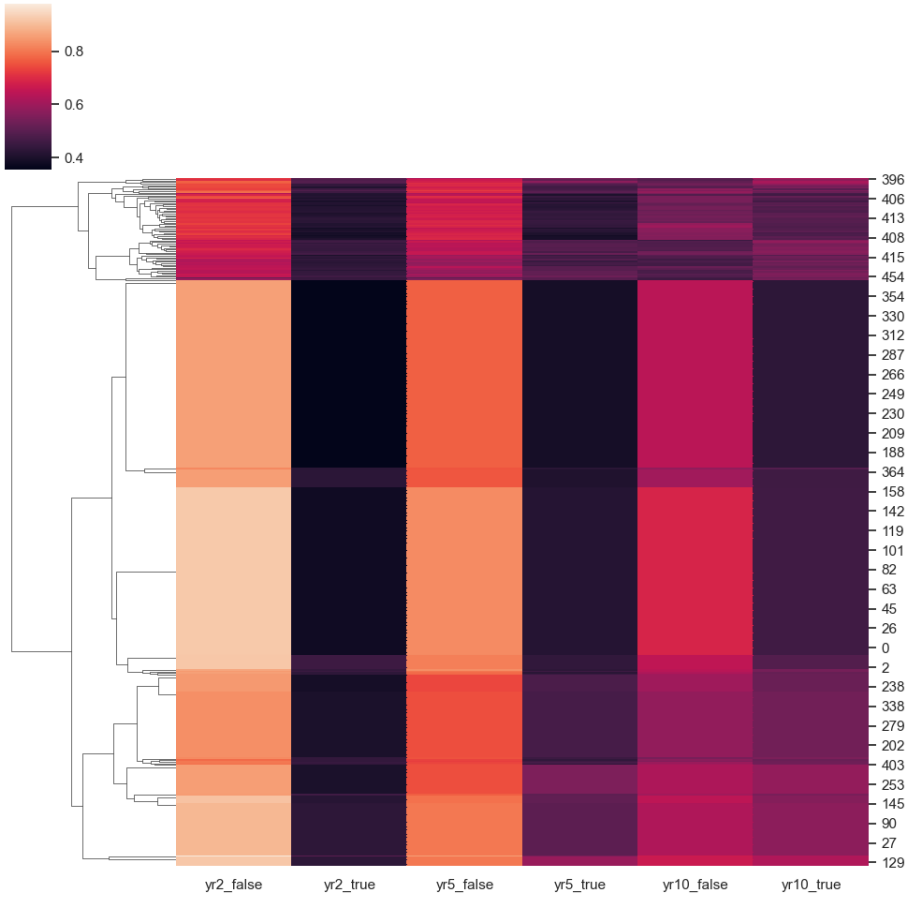

Figure 4.2. Raw visualization of CKD progression risk. The label in each row represents the index of the patient, and the shade of the color represents the confidence level of the label.

Figure 4.3 illustrates the true/false ratio of each class. The value $\log_2(\frac{true}{false})$ is positive if the number of patients predicted to be true is larger than the number of false, and negative otherwise. It is clear that the majority is negative except a part of class year10.

It is even clearer in Figure 4.4 which is binary, i.e. 1 if the possibility of true is greater than that of false, 0 otherwise. As shown in Figure 4.4, the majority is black which means 0.

Figure 4.5 shows predictions made by the DVAE model the random forest classifier. They both tended to predict all inputs to a single class which still affected by the imbalanced dataset. Figure 4.6 gives out their classification reports. As mentioned in the precious chapters, although the random forest model outperformed the DVAE model, people cannot fully trust it because it is not interpretable.

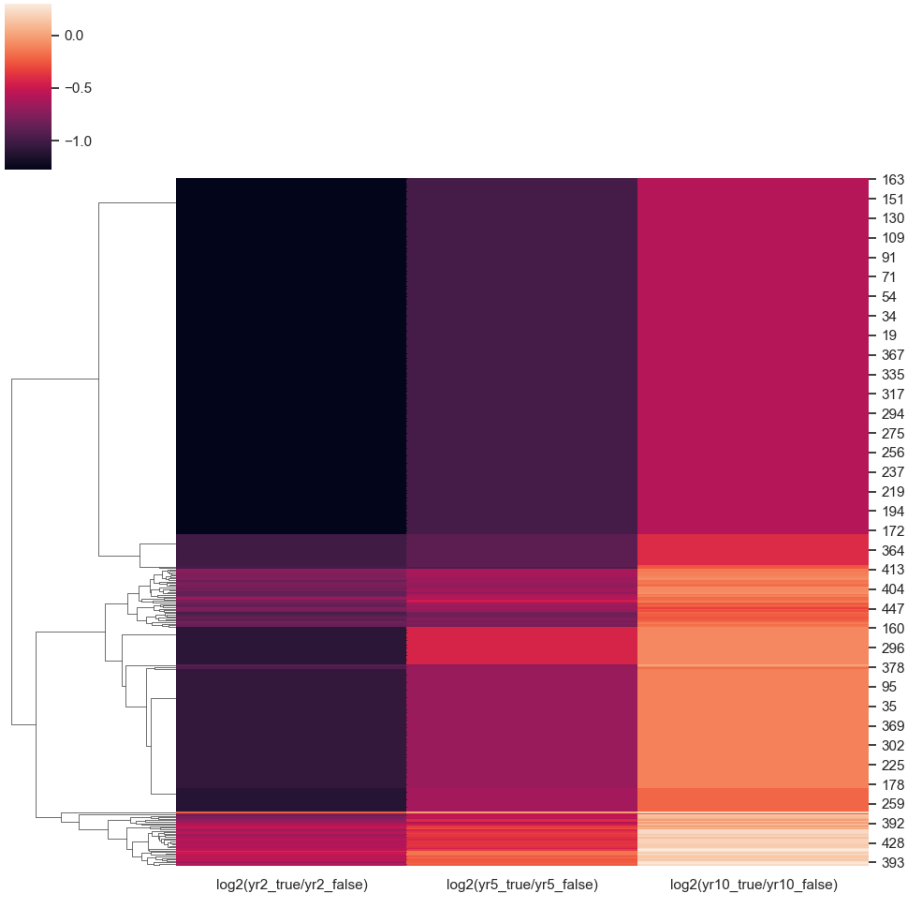

Figure 4.3. Prediction of CKD progression risk. The label in each row represents the index of the patient, and the shade of the color represents the confidence level of the label.

4.2.2 Interpretability

As mentioned in previous chapters, the interpretability is of great significance in the medical field. The DVAE model has inherent advantages being interpretable. As described in Figure 2.1, the input data can be encoded to some latent space features that reveal the hidden relationship among clinical variables. As shown in Figure 4.7, given the input graph x containing the a patient and their states, the encoder compressed it into the latent space. Vectors in the latent space had multiple dimensions and each dimension was expected to represent some information interpretable to human beings.

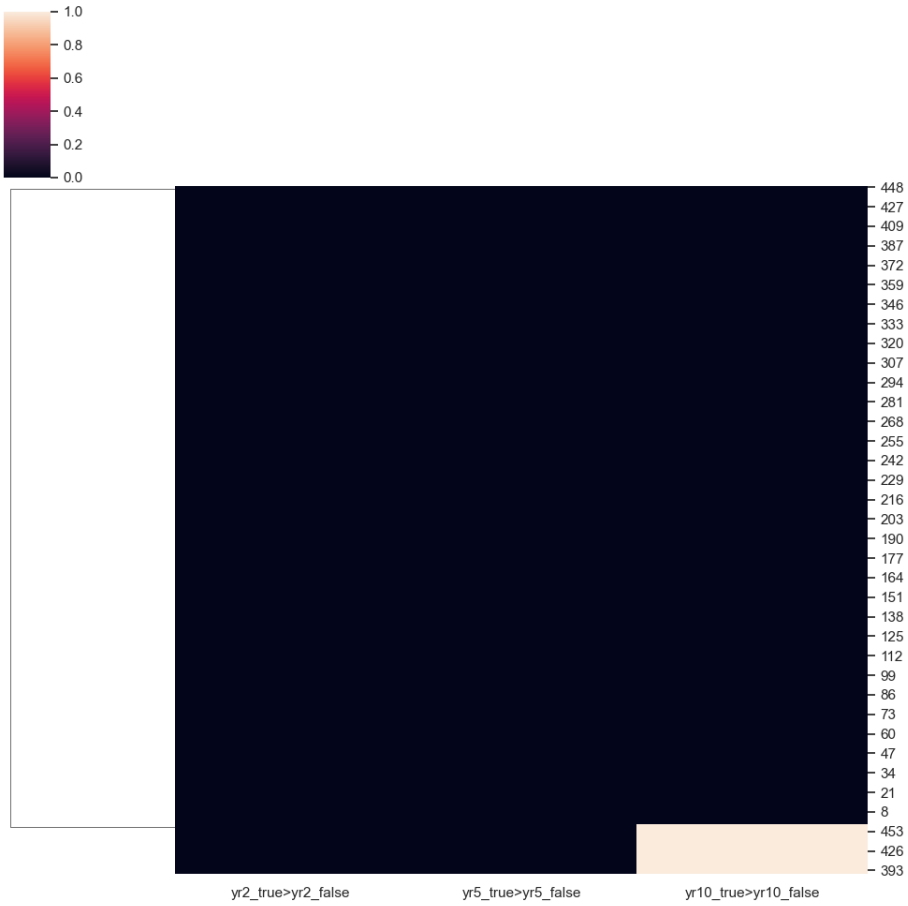

Figure 4.4. Prediction of CKD progression. The label in each row represents the index of the patient, and the shade of the color represents the confidence level of the label. The labels are binary, so any value is either 0 or 1.

The researcher gradually changed one dimension in the latent space. Specifically, the researcher set 5 different values from the minimal to the maximal and fixed all other dimensions. Then took these 5 feature vectors as input, and used the decoder to reconstruct the relationship graph from the patient to states. Thus, 5 reconstructed graphs were generated. The researcher repeated the process on 2 more dimensions.

Figure 4.7 shows the results, each figure representing the results of changing one dimension. Taking the figure on the top as an instance, predicted confidence levels for each state were grouped together. The abscissa represents the indices of the feature, and the ordinate represents the predicted confidence levels. Most features were affected little, e.g. the 18th variable CHOLESTEROL_2, while some features were affected significantly, e.g. the 22nd variable

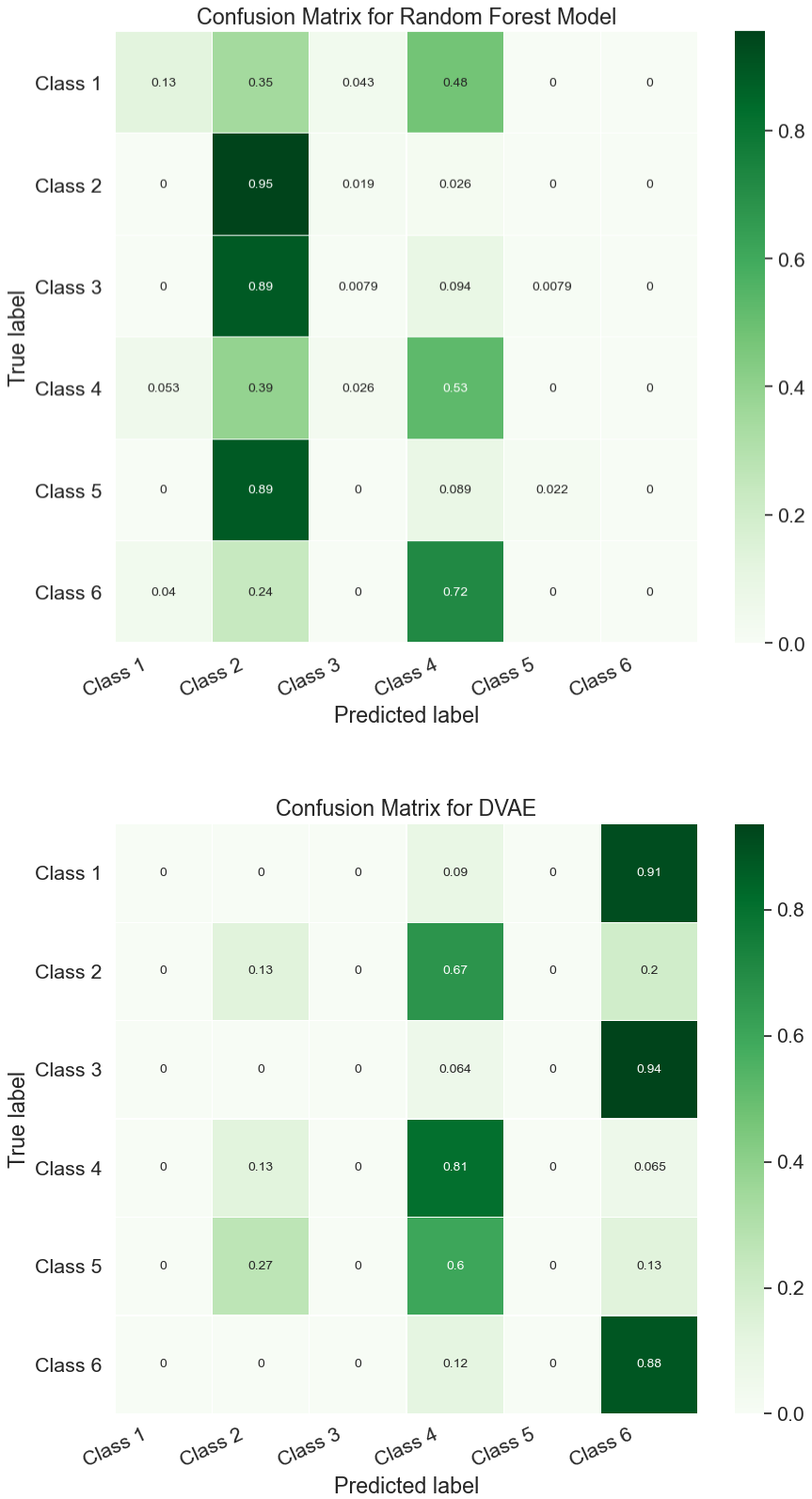

Figure 4.5. Confusion matrix

	DVAE				Random forest			
	precision	recall	f1-score	support	precision	recall	f1-score	support
0	0.00	0.00	0.00	134	0.50	0.13	0.21	23
1	0.23	0.13	0.17	45	0.40	0.95	0.56	155
2	0.00	0.00	0.00	110	0.17	0.01	0.02	127
3	0.21	0.81	0.33	31	0.27	0.53	0.36	38
4	0.00	0.00	0.00	60	0.67	0.02	0.04	90
5	0.22	0.88	0.35	78	0.00	0.00	0.00	25
accuracy			0.22	458			0.38	458
macro avg	0.11	0.30	0.14	458	0.33	0.27	0.20	458
weighted avg	0.07	0.22	0.10	458	0.36	0.38	0.24	458

Figure 4.6. Classification reports

CHOLESTEROL_2. Besides, the 5 corresponding predictions on the CKD event occurring time were: (1) less than 2 years; (2) 2 years to 2-5 years; (3) 5-10 years; (4) 2-5 years; (5) 2-5 years. This first increasing and then decreasing trend happened to be the opposite of the 22nd variable which first decreased and then increased, which might imply they have the hidden relationship.

Similar relationships appeared as well in the other two figures in Figure 4.7.

4.3 Summary

Empirical results are given in this chapter. The DVAE model is compared with the random forest classifier and apparently outperforms it. In addition, the interpretability of the DVAE model is also discussed in this chapter.

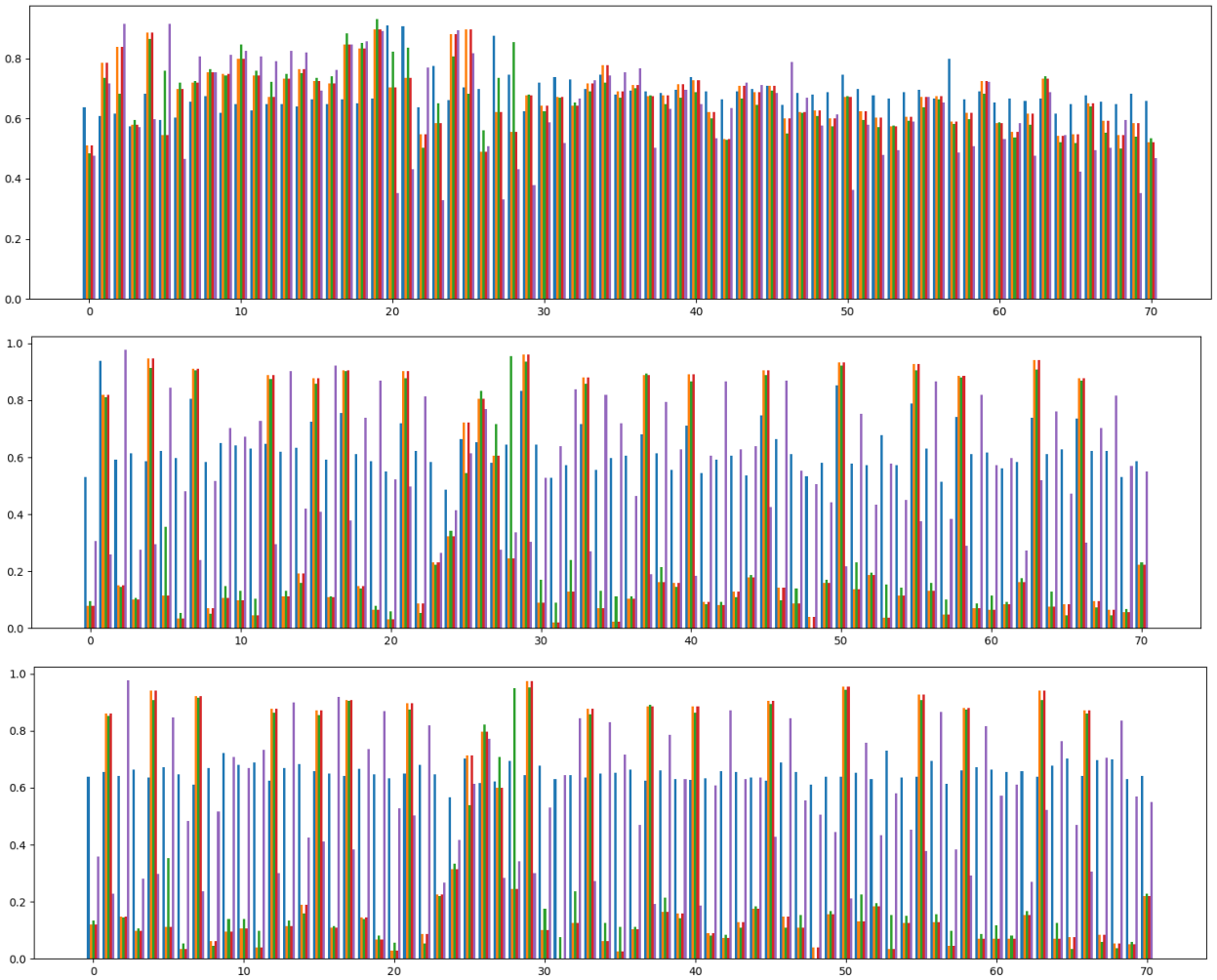

Figure 4.7. Reconstructed representations. The abscissa represents the indices of the feature, and the ordinate represents the predicted confidence levels.

CHAPTER 5. SUMMARY AND FUTURE PLAN

5.1 Summary

In this paper, the problem of disentangled representations learning was studied. The researcher adopted a disentangled representation learning model in the medical field, building a DVAE model that predicts the CKD progression as well as reveals the relationship between clinical variables and the CKD progression. The DVAE model was compared with a baseline model in terms of the evaluation criteria.

Chapter 1 gave an introduction of the research question. It introduced the background of the problem, states the issues existing in the current situation, and explained the research significance, assumptions, limitations, and delimitations.

Chapter 2 reviewed the related research works, providing the necessary background theologies. First, it reviewed recent works studying the COVID-19 sequelae. These works proved that the problem studied in this project indeed existed. In addition, state-of-art machine learning models used in the medical field were also reviewed in this chapter. However, they were not interpretable which was the research gap filled up in this study.

Chapter 3 described the proposed solution. It first presented the dataset used in experiments. Second, it introduced the data preprocessing procedure. Third, It described the structure of the proposed model and the criteria used to evaluate the model. Additionally, in order to handle the imbalanced data, two methods were introduced in this chapter.

Chapter 4 evaluated the proposed approach and compares it with the baseline model. First, it introduced the experimental setup and the baseline model as well as the evaluation metrics. Second, it presented the experimental results including the accuracy, the cross entropy loss, training logs, the visualization of CKD progression risk and the running time. Besides, it discussed the interpretability of the DVAE model, proving that the DVAE model outperformed the baseline model.

Chapter 5 summarized the thesis and discussed plans for the future work.

In terms of the research questions proposed in Chapter 1, they now can be answer here:

1. The proposed method based on a variational autoencoder took clinical variables as inputs and predicted CKD progression risk as described in previous chapters.

2. As shown in Chapter 3, changing one dimension in the encoded vector resulted in some values changing and some remaining same. Certainly, those clinical variables might have hidden relationships.

3. As shown in Chapter 3, changing one dimension in the encoded vector resulted in predictions changing as well. Besides, they shared a similar pattern with some clinical variables. On the other hand, it can be understood as: if those variables change in this way, the prediction will be as above.

5.2 Future Plan

The future plan mainly includes the following two aspects:

1. As mentioned earlier, because the dataset is imbalanced, the empirical results are not good enough for some classes. This paper has tried adding weights to balance, but the effect is not very significant. The researcher considers exploring other methods in the future or trying to obtain a more ideal dataset.

2. In Section 4.2.2, the researcher has demonstrated that gradually changing a dimension in the latent space will affect some features in the reconstruction. Besides, some patterns are observed. The researchers plan to take advantage of medical knowledge in the future to explain the meaning of these changes and their impact on patients.

REFERENCES

Alimadadi, A., for Hypertension, C., Medicine, P., Aryal, S., Manandhar, I., Munroe, P. B., ... et al. (2020, Apr). *Artificial intelligence and machine learning to fight covid-19.* Retrieved from https://journals.physiology.org/doi/10.1152/physiolgenomics.00029.2020

Barker-Davies, R. M., O'Sullivan, O., Senaratne, K. P., Baker, P., Cranley, M., Dharm-Datta, S., ... et al. (2020). The stanford hall consensus statement for post-covid-19 rehabilitation. *British Journal of Sports Medicine*, *54*(16), 949–959. doi: 10.1136/bjsports-2020-102596

Bengio, Y., Courville, A. C., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, *35*(8), 1798–1828. Retrieved from https://doi.org/10.1109/TPAMI.2013.50 doi: 10.1109/TPAMI.2013.50

Bennett, J., Lanning, S., et al. (2007). The netflix prize. In *Proceedings of kdd cup and workshop* (Vol. 2007, p. 35).

Berenguer, A. D., Sahli, H., Joukovsky, B., Kvasnytsia, M., Dirks, I., Alioscha-Perez, M., ... et al. (2020, Nov). *Explainable-by-design semi-supervised representation learning for covid-19 diagnosis from ct imaging.* Retrieved from https://arxiv.org/abs/2011.11719v1

Booth, A. L., Abels, E., & McCaffrey, P. (2020). Development of a prognostic model for mortality in covid-19 infection using machine learning. *Modern Pathology*, *34*(3), 522–531. doi: 10.1038/s41379-020-00700-x

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In L. Cao, C. Zhang, T. Joachims, G. I. Webb, D. D. Margineantu, & G. Williams (Eds.), *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, sydney, nsw, australia, august 10-13, 2015* (pp. 1721–1730). ACM. Retrieved from https://doi.org/10.1145/2783258.2788613 doi: 10.1145/2783258.2788613

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8). Retrieved from https://www.mdpi.com/2079-9292/8/8/832 doi: 10.3390/electronics8080832

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems 29: Annual conference on neural information processing systems 2016, december 5-10, 2016, barcelona, spain* (pp. 2172–2180). Retrieved from https://proceedings.neurips.cc/paper/2016/hash/7c9d0b1f96aebd7b5eca8c3edaa19ebb-Abstract.html

Covid-19 map. (n.d.). Retrieved from https://coronavirus.jhu.edu/map.html

Dawei Wang, M. (2020, Mar). *Clinical characteristics of patients with 2019 novel coronavirus (2019-ncov)–infected pneumonia in wuhan, china.* JAMA Network. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2761044

Demertzis, Z. D., Dagher, C., Malette, K. M., Fadel, R. A., Bradley, P. B., Brar, I., ... Suleyman, G. (2020). Cardiac sequelae of novel coronavirus disease 2019 (covid-19): a clinical case series. *European Heart Journal - Case Reports*, *4*(FI1), 1–6. doi: 10.1093/ehjcr/ytaa179

Doersch, C. (2016). Tutorial on variational autoencoders. *CoRR*, *abs/1606.05908*. Retrieved from http://arxiv.org/abs/1606.05908

Du, M., Liu, N., & Hu, X. (2020). Techniques for interpretable machine learning. *Commun. ACM*, *63*(1), 68–77. Retrieved from https://doi.org/10.1145/3359786 doi: 10.1145/3359786

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In F. Bonchi, F. J. Provost, T. Eliassi-Rad, W. Wang, C. Cattuto, & R. Ghani (Eds.), *5th IEEE international conference on data science and advanced analytics, DSAA 2018, turin, italy, october 1-3, 2018* (pp. 80–89). IEEE. Retrieved from https://doi.org/10.1109/DSAA.2018.00018 doi: 10.1109/DSAA.2018.00018

Guan, W.-j., Al., E., for the China Medical Treatment Expert Group for Covid-19*, the State Key Laboratory of Respiratory Disease, A. A., Shimabukuro, T. T., Others, ... Others, J. S. a. (2020, May). *Clinical characteristics of coronavirus disease 2019 in china: Nejm.* Retrieved from https://www.nejm.org/doi/full/10.1056/NEJMoa2002032

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, *5*(4), 1–19.

Herrera-Valdés, R., Almaguer-López, M., López-Marín, L., Bacallao-Méndez, R., Guerra-Bustillo, G., et al. (2020). Covid-19 and the kidneys: Risk, damage and sequelae. *MEDICC review*, *22*(4), 87–88.

Janiri, D., Kotzalidis, G. D., Giuseppin, G., Molinaro, M., Modica, M., Montanari, S., ... et al. (2020, Oct). *Psychological distress after covid-19 recovery: Reciprocal effects with temperament and emotional dysregulation. an exploratory study of patients over 60 years of age assessed in a post-acute care service.* Frontiers. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyt.2020.590135/full

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. In Y. Bengio & Y. LeCun (Eds.), *2nd international conference on learning representations, ICLR 2014, banff, ab, canada, april 14-16, 2014, conference track proceedings.* Retrieved from http://arxiv.org/abs/1312.6114

Lai, C.-C., Ko, W.-C., Lee, P.-I., Jean, S.-S., & Hsueh, P.-R. (2020). Extra-respiratory manifestations of covid-19. *International journal of antimicrobial agents*, *56*(2), 106024.

Lam, M. H.-B. (2009). Mental morbidities and chronic fatigue in severe acute respiratory syndrome survivors. *Archives of Internal Medicine*, *169*(22), 2142. doi: 10.1001/archinternmed.2009.384

Liang, D., Krishnan, R. G., Hoffman, M. D., & Jebara, T. (2018). Variational autoencoders for collaborative filtering. In P. Champin, F. Gandon, M. Lalmas, & P. G. Ipeirotis (Eds.), *Proceedings of the 2018 world wide web conference on world wide web, WWW 2018, lyon, france, april 23-27, 2018* (pp. 689–698). ACM. Retrieved from https://doi.org/10.1145/3178876.3186150 doi: 10.1145/3178876.3186150

Locatello, F., Bauer, S., Lucic, M., Rätsch, G., Gelly, S., Schölkopf, B., & Bachem, O. (2019). Challenging common assumptions in the unsupervised learning of disentangled representations. In *Reproducibility in machine learning, ICLR 2019 workshop, new orleans, louisiana, united states, may 6, 2019.* OpenReview.net. Retrieved from https://openreview.net/forum?id=Byg6VhUp8V

Ma, J., Zhou, C., Cui, P., Yang, H., & Zhu, W. (2019, Oct). *Learning disentangled representations for recommendation.* Retrieved from https://arxiv.org/abs/1910.14238v1

Marlin, B. (2004). *Collaborative filtering: A machine learning perspective*. University of Toronto Toronto.

Martinez-Rojas, M. A., Unit, M. P., Vega-Vega, O., Nephrology, D. o., Bobadilla, N. A., SK, A., ... et al. (2020, Jun). *Is the kidney a target of sars-cov-2?* Retrieved from https://journals.physiology.org/doi/full/10.1152/ajprenal.00160.2020

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, *267*, 1–38. Retrieved from https://doi.org/10.1016/j.artint.2018.07.007 doi: 10.1016/j.artint.2018.07.007

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications. *CoRR*, *abs/1901.04592*. Retrieved from http://arxiv.org/abs/1901.04592

Nguyen, A. M., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE conference on computer vision and pattern recognition, CVPR 2015, boston, ma, usa, june 7-12, 2015* (pp. 427–436). IEEE Computer Society. Retrieved from https://doi.org/10.1109/CVPR.2015.7298640 doi: 10.1109/CVPR.2015.7298640

Patel, D., Kher, V., Desai, B., Lei, X., Cen, S., Nanda, N., ... et al. (2021, Feb). *Machine learning based predictors for covid-19 disease severity.* Nature Publishing Group. Retrieved from https://www.nature.com/articles/s41598-021-83967-7

Polino, A., Pascanu, R., & Alistarh, D. (2018). Model compression via distillation and quantization. In *6th international conference on learning representations, ICLR 2018, vancouver, bc, canada, april 30 - may 3, 2018, conference track proceedings.* OpenReview.net. Retrieved from https://openreview.net/forum?id=S1XolQbRW

Provost, F. (2000). Machine learning from imbalanced data sets 101. In *Proceedings of the aaai'2000 workshop on imbalanced data sets* (Vol. 68, pp. 1–3).

Rogers, J. P., Chesney, E., Oliver, D., Pollak, T. A., McGuire, P., Fusar-Poli, P., ... David, A. S. (2020). Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the covid-19 pandemic. *The Lancet Psychiatry*, *7*(7), 611–627. doi: 10.1016/s2215-0366(20)30203-0

Vaira, L. A., Salzano, G., Deiana, G., & De Riu, G. (2020). Anosmia and ageusia: common findings in covid-19 patients. *The Laryngoscope*, *130*(7), 1787–1787.

Wang, L., Li, X., Chen, H., Yan, S., Li, Y., Li, D., & Gong, Z. (2020). Sars-cov-2 infection does not significantly cause acute renal injury: An analysis of 116 hospitalized patients with covid-19 in a single hospital, wuhan, china. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3541116

Wu, M., Hughes, M. C., Parbhoo, S., Zazzi, M., Roth, V., & Doshi-Velez, F. (2018). Beyond sparsity: Tree regularization of deep models for interpretability. In S. A. McIlraith & K. Q. Weinberger (Eds.), *Proceedings of the thirty-second AAAI conference on artificial intelligence, (aaai-18), the 30th innovative applications of artificial intelligence (iaai-18), and the 8th AAAI symposium on educational advances in artificial intelligence (eaai-18), new orleans, louisiana, usa, february 2-7, 2018* (pp. 1670–1678). AAAI Press. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16285

Xiao, L.-s., Li, P., Sun, F., Zhang, Y., Xu, C., Zhu, H., … et al. (2020). Development and validation of a deep learning-based model using computed tomography imaging for predicting disease severity of coronavirus disease 2019. *Frontiers in Bioengineering and Biotechnology*, *8*. doi: 10.3389/fbioe.2020.00898