Early Identification of Pulmonary Embolism In Intensive Care Unit

Timothy Green
MU Informatics Institute
University of Missouri
Columbia MO, USA
greentim@missouri.edu

Peng Zhao
MU Informatics Institute
University of Missouri
Columbia MO, USA
pz72b@mail.missouri.edu

Yuanyuan Shen
MU Informatics Institute
University of Missouri
Columbia MO, USA
yshpf@mail.missouri.edu

Yan Zhuang
MU Informatics Institute
University of Missouri
Columbia MO, USA
yznm9@mail.missouri.edu

Abstract—Pulmonary Embolism(PE), can be life-threatening without rapid appropriate therapy and often leads to chronic disease, disability, and increasing health care costs. The ambiguity of symptoms makes PE difficult to diagnose, and available imaging strategies have their limitations. Still, it is regarded as one of the most preventable of the major complications. Morbidity and mortality caused by this disease have not essentially changed in the past twenty years despite implementation of the national prophylaxis guidelines. Current prophylaxis measures may be insufficient. In this article, we aim to apply some data mining methods to predict early onset of PE in intensive care unit patients using lab tests and electrocardiography(ECG). Early warning of risk for PE could prompt earlier prophylactic procedures, leading to improved outcomes. Data were obtained from MIMIC III, a relational database containing tables of data relating to patients who stayed within the intensive care units at Beth Israel Deaconess Medical Center, and the MIMIC III Matched Waveform dataset, containing high resolution ECG waveform data during those stays.

Keywords: Pulmonary Embolism, Intensive care unit(ICU), electrocardiography(ECG), electronic health record (EHR).

INTRODUCTION

Measured by morbidity and mortality, very few health issues could compete with deep vein thrombosis (DVT) and pulmonary embolism (PE), generally known as venous thromboembolism (VTE) [1]–[4]. It affects approximately 100 per 100,000 person-years and leads to 200,000–300,000 hospitalizations yearly in the United States [1], [5]–[7]. Venous thromboembolism (VTE), including deep vein thrombosis (DVT) and PE, results in approximately 100,000 deaths annually in the United States, of which acute PE is responsible for 80,000 [1], [7]–[9]. Further, given that PE may be a common cause of mortality outside of the hospital, the true rate of PE development may be underestimated [10]. As such, VTE is the third most common cause of cardiovascular death, after myocardial infarction and stroke [1]. The high prevalence of PE and the proportion of patients requiring intensive care result in a substantial cost burden to these patients and society. The cost per hospitalization for a primary PE diagnosis ranges between $8,000 and $30,000 and is even higher for rehospitalization or recurrent PE [11], [12].

The diagnosis of PE is composed of commonly seen diseases. The symptoms and findings are nonspecific, and clinical diagnosis is likely not reliable. Therefore, the PE could be overlooked because of the conditions of primary diseases, and the diagnosis can be easily delayed. Lately, some state of the art developments have been introduced to the diagnosis and treatment of PE. But the gold standard approach and procedure for PE is still not available.

The ACCP 9th edition guideline has been published to make consensus on the diagnosis, treatment and prophylaxis of PE, which requires multidisciplinary cooperation [13]–[15]. Pulmonary angiography is one of the definitive diagnostic methods for PE, but it is invasive and expensive [16], [17]. Rather than these invasive ways, noninvasive diagnostic methods are always recommended and preferred, such as ultrasonography (USG), ECG along with a combination of various clinical and laboratory findings.

The objective of this article is to use data mining to retrieve and analyze ECG waveform and clinical data to aid in the diagnosis of PE in multiple ICU patients. Such a diagnostic algorithm could help clinicians to decide the best time window with which to proceed with prophylactic procedures, and reduce morbidity, mortality, and healthcare costs. ECG is a noninvasive and very useful test, also widely used in multiple clinical practice for the diagnosis of some certain diseases that involve the cardiovascular disease. The PE increases the pulmonary arterial pressure in the proportion with the level of anatomic pulmonary vascular obstruction and exposes right ventricle to strain that could be reflected by the ECG.

In this project, we developed a PE incidence diagnostic model based on the critical care data with the KDD process. Both descriptive and predictive data analysis were performed with data mining models. In the descriptive data analysis, we determined (1) clinical variables frequently associated with the diagnosis of PE, and (2) trends of the ECG time series data strongly associated with the diagnosis of PE. The outcome of the descriptive data analysis was used to guide the predictive model development (e.g., variable selection).

I. Knowledge Discovery and Data Mining

A. Data Selection

The MIMIC III dataset [18] and MIMIC III Waveform Database Matched Subset [19] are freely available de-identified clinical datasets comprising data on up to 40,000 critical care patients seen at Beth Israel Deaconess Hospital between 2001 and 2012. This is an unprecedented health care research dataset because it contains matched waveform comprised of waveform readings of various types, including ECG Lead II at 125 Hz, as well as 1 Hz heart and respiration rates, SpO2, and blood pressure readings. Many clinical decision support (CDS) rules have been developed using point in time lab and clinical readings, however very high-resolution streaming clinical data such as ECG waveforms presents an opportunity to create predictive decision support that can be triggered much sooner and at potentially lower cost, as expensive lab tests could be bypassed. For these reasons, the MIMIC III dataset was chosen to create a novel diagnostic algorithm addressing PE development.

The MIMIC III data contains 58,976 transactions. All patients' records are from ICU stays. There are 22,317 waveforms records across 10,282 clinical records. Our data selection process identified 1,061 PE patients, with 327 of those patients having ECG information (the PE class). We have pre-selected eight lab tests attributes based on medical domain knowledge. All PE patients would have over 500 ng/mL in D-dimer test [20]. But not all patients with high D-dimer test values would be diagnosed as PE. We have also chosen seven other tests which potentially related to PE as Blood gas test (PaO2), White blood cells count (WBC), Erythrocyte sedimentation rate (ESR), Lactate dehydrogenase (LDH), creatine phosphokinase test (CPK), Aspartate transaminase (AST) and Fibrin degradation products (FDPS). These attributes would be classified as normal or abnormal according to the standard ranges from medical textbook. In the MIMIC III dataset, several lab tests have been recorded into multiple attributes although they represent the same lab test. We have converted those values using the same standard and merged them into one attribute to maintain a list of seven total attributes.

There are 57,867 non-PE patients (the non-PE class). Performing machine learning on such unbalanced classes can have detrimental effects on the results. To eliminate the effects of a highly skewed data set, a matched subset of non-PE patients was derived from the non-PE patients that have MIMIC III ECG waveforms. The control group consists of 350 patients. Figure 1 depicts the selection criteria for the study and control groups. The control group has a similar distribution of demographic attributes as race and gender. Table 1 describes the demographics of each class. Because the D-dimer test has a significant impact on PE diagnosis, there are no PE patients with a D-dimer test with less than 500 ng/mL. In order to determine what other factors are involved in patients diagnosed with PE in addition to an abnormal D-dimer test, all patients in the control group also have over 500 ng/mL in D-dimer test.

Table 1 Demographics of each class

Attributes	Amount
American Indian or Alaska Native	2 (0.57%)
Asian	5 (1.42%)
Black or African American	35 (9.97%)
Native Hawaiian or Other Pacific Islander	20 (5.70%)
White	250 (71.23%)
Other	38 (10.83%)
Female	169 (48.14%)
Male	181 (51.86%)

In the waveform data, each instance of a patient being connected to ECG is stored as a *segment*. Lead II ECG data was selected for analysis due to its prevalence in the dataset and the common nature of collection of Lead II as part of the standard of care. For a given patient, there can be zero or more ECG segments with Lead II data. In the PE class, every patient had at least one Lead II segment. Segments shorter than 5 minutes were not used in this analysis as the ECG data were found to be of poor quality. After cleansing the data (Figure 1), 202 PE patients and 176 non-PE patient remained in the dataset.

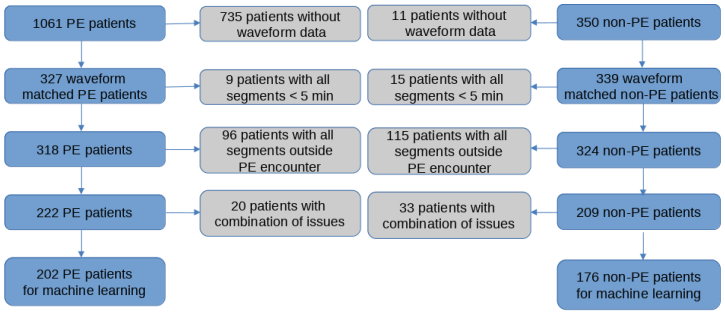

Figure 1 – PE and non-PE class data selection for ECG waveform data

B. Data Preprocessing and Transformation

We will itemize the attributes values, as age group, weight, etc. which will be used for frequent pattern mining.

ECG waveform readings are stored in a binary format, at a frequency of 125 Hz. In this study, we used the ECG waveform from Lead II, which was present for every patient in our dataset. Using the algorithm from [21], the start and end of the QRS complexes in the ECG waveforms were identified. Q, R, and S peak values were then identified using the algorithm in Figure 2. Figure 3 shows typical identified QRS complexes with Q, R, and S peaks marked.

```
For all qrs_start,qrs_end in QRS_complexes
  QRS_complex ← all waveform data points from qrs_start to qrs_end
  r_index,r_value ← max(QRS_complex)
  if r_index == qrs_start
```

```
  then q_value = None
  else q_value = min(QRS_complex(qrs_start:r_index))
if r_index == qrs_end
  then s_value = None
  else s_value = min(QRS_complex(r_index:qrs_end))
```

Figure 2 – Q, R, S peak detection

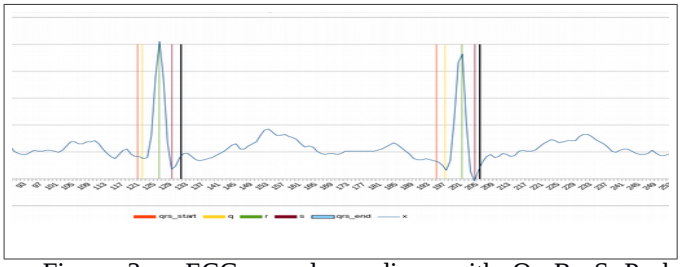

Figure 3 – ECG sample readings with Q, R, S Peak Detection

The resulting marked Q, R, and S peaks were output as sequences for further analysis by neural network processing. In addition to peak data, the duration of each QRS complex in milliseconds was output as a matched sequence to the Q, R, and S peak data.

There are also many missing values in the chart data. MIMIC III contains not only recorded heart rate, but also blood pressure and lab test values with time stamps. Imputation of certain missing clinical values may be possible using other correlated clinical data as well. Missing data unable to imputed would be left out of the dataset. A determination would be made as to whether missing data would require the complete removal of the patient record. Noisy data is expected to be a problem especially within the ECG waveform data. Because of the high resolution of the data, it may be possible to determine outliers and completely remove the noisy data from the ECG set. Evaluation of noise will need to be made once the full dataset is available for processing.

C. Data Mining

ECG Signal Time Series Classification

In recent years, deep learning has emerged as a popular technique in time series study because it requires no feature engineering compared to traditional time series analysis methods. Recurrent neural network (RNN) is a type of artificial neural network that is capable of classifying sequence data with temporal dependence [22]. As a variant of RNN, long short-term memory (LSTM) network addresses the traditional RNN's exploding and vanishing gradient problems and is relatively less sensitive to the length of gaps [23].

In this work, we have developed a supervised multivariate time series classification model with the time series data of Q, R, S peaks, and complex durations based on LSTM. This is a binary classification task to classify if a patient will develop PE or not based on sequences with 4 features. Unlike the traditional feed-forward neural networks, the input data of LSTM is three-dimensional with one additional temporal dimension.

Table 2. Class distribution in the training and testing sets.

	Training	Testing
PE	133	58
Non-PE	133	34
Total	266	92

After removing patients with sequences shorter than 5000, the remaining 358 patients were split into a training set and a testing set. Table 2 details the class distribution in these two sets. Rectified linear unit (ReLU) [24] and softmax function [25] were used to activate the hidden layer(s) and the output layer, respectively. Cross entropy loss and accuracy were used to monitor the training and testing progresses. Each of the candidate models has been trained for a total of 300 epochs on the training set in batch mode with 50 patients' data in each batch iteration. At the end of each training iteration, the model's performance was tested using the testing set. The learning rate and lambda loss were set to 0.002 and 0.001, respectively. Each hidden layer had 10 neurons. The training was performed on the TensorFlow framework [26].

(1) Effect of the sequence length

The length of a sequence (L, the number of time steps) is a key determinant of the LSTM model's performance. If the sequence is too short, the characteristic temporal patterns contrasting PE versus non-PE may not occur. If it is too long, the model may "forget" the characteristic temporal patterns due to increased noises. Figures 4-6 shows the effect of sequence length on the model's performance on the testing set when there was 1 hidden layer. It can be seen that when the first 500 time steps were used, the model achieved an accuracy of 0.48 on the testing set at the end of the training. When the sequence length was increased to 1000, the accuracy became into 0.60. If the sequence length was further increased to 2000, the model became unstable and the accuracy curve had a decreasing trend with an ending accuracy of 0.59.

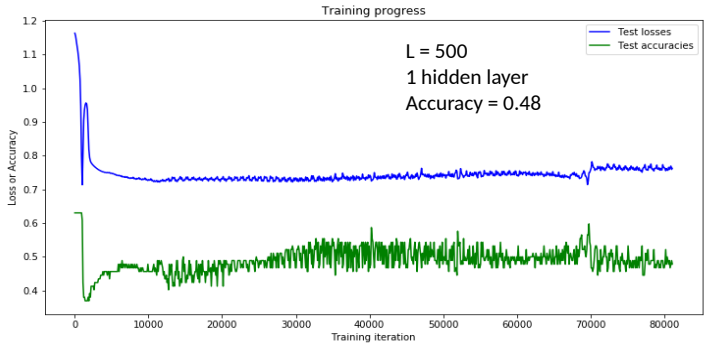

Figure 4. Model performance on the testing set during the training progress (L = 500, 1 hidden layer).

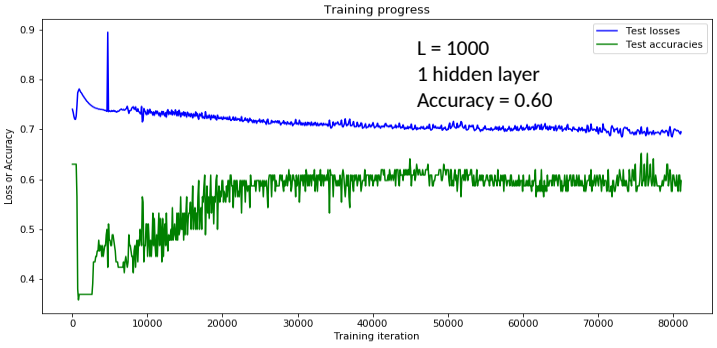

Figure 5. Model performance on the testing set during the training progress (L = 1000, 1 hidden layer).

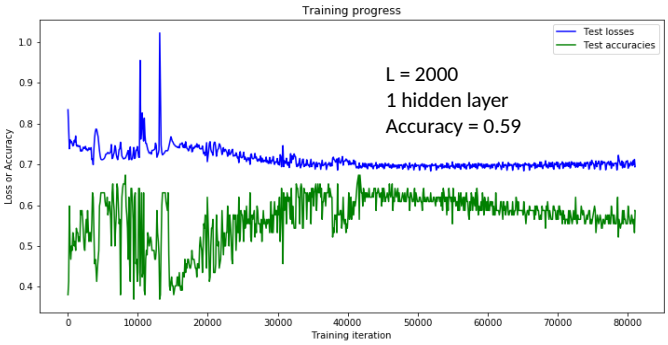

Figure 6. Model performance on the testing set during the training progress (L = 2000, 1 hidden layer).

(2) Effect of the hidden layer number

A deeper model can achieve a higher level of abstraction of the data. The number of hidden layers of the LSTM model is another important hyper-parameter that needs to be tuned. In this work, the deep LSTM models were constructed by stacking basic LSTM cells. Figure 7-8 show the effect of the hidden layer number on the model's performance on the testing set when the sequence length was limited to 1000. It can be seen that when there were 2 hidden layers, the accuracy curve became unstable and reached 0.60 at the end of the training. When the hidden layer number was increased to 3, the accuracy curve became much noisier and the accuracy dropped to 0.49.

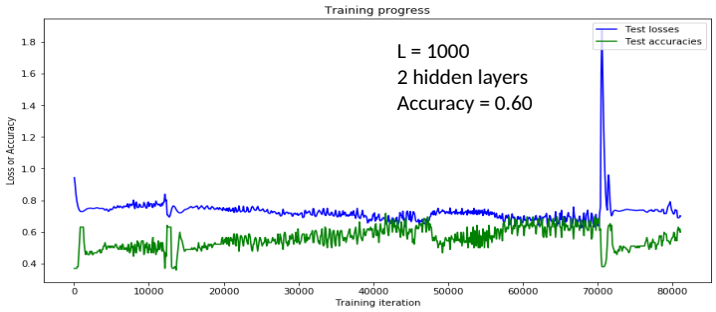

Figure 7. Model performance on the testing set during the training progress (L = 1000, 2 hidden layers).

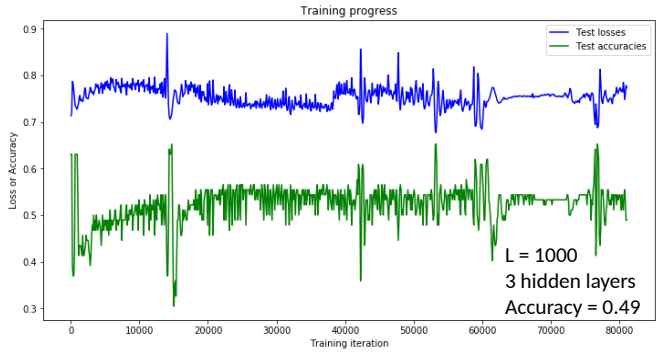

Figure 8. Model performance on the testing set during the training progress (L = 1000, 3 hidden layers).

The model trained with the first 1000 time steps using 1 hidden layer had the best accuracy (0.60). In this work, the data of 266 training patients were insufficient for building a deep model. Therefore, increasing the number of hidden layer did not help to improve the accuracy. In addition, the sequence was not the longer the better. This can be influenced by the inherent temporal pattern of the sequences and also the number of training samples.

Contrast Mining

Association Rule Mining (ARM) is a popular algorithm using pattern discovery [27]. ARM can find patterns (a group of attributes) that appear at a high frequency. ARM works good on sparse dataset. As long as the threshold is defined properly, user could filter out all frequent patterns quickly. Due to the difficulty of inferring PE diagnosed time, and since the list of pre-selected lab tests are non-specific, we need feature selection to extract those attributes worth further study. We used contrast set mining (CSM) to perform this task[28]. CSM is a form of ARM but can find all patterns strongly indicative of one class over another.

Support is represented as frequency of a pattern. After a user defines a frequency threshold, any pattern that passes the minimum support threshold would be considered as frequent pattern. Growth is represented as the ratio between supports values of the same pattern from different classes. The growth rate is an important feature to evaluate the performance of a pattern, and is easily interpreted.

We have selected 350 PE patients as study group and selected 350 non-PE patients with the same demographic distributions. We kept demographic attributes and all abnormal lab tests for each single transaction. After performing contrast set mining and collecting all highly contrasted patterns, we used logistic regression on those patterns to test how well the patterns correlated with PE diagnosis. We used the Chi-square test to measure the fit of the model. In this case, statistical significance of a pattern is more meaningful than a single attribute. All of the patterns whose growth are over 2 along with support, growth and p-value are listed in Table 3.

Table 3. Highly contrasted patterns

Pattern	Support	Growth	P-Value
olderthan60, Blood Gas, Sedimentation Rate	0.0314	3.6667	< 0.0001
F, White Blood Cell, D-Dimer	0.0429	2.1429	< 0.0001
DIVORCED, Blood Gas, NO	0.0400	2.8000	0.0001
SINGLE, Blood Gas, Sedimentation Rate	0.0257	2.2500	0.0017
lessthan40, Blood Gas, F	0.0343	2.4000	< 0.0001
SINGLE, F, Sedimentation Rate	0.0200	7.0000	0.4156
DIVORCED, Blood Gas	0.0457	2.2857	0.0007
Sedimentation Rate, Blood Gas	0.0600	2.1000	0.0005
NO, Blood Gas, WIDOWED	0.0457	2.2857	0.0001
DIVORCED, Blood Gas, WHITE	0.0343	2.4000	0.0023
Sedimentation Rate, White Blood Cell, Blood Gas	0.0543	2.1111	< 0.0001
Sedimentation Rate, MARRIED, M	0.0343	3.0000	0.1444
olderthan60, Sedimentation Rate, M	0.0257	2.2500	< 0.0001
lessthan60, Sedimentation Rate, M	0.0200	2.3333	< 0.0001
M, HISPANIC OR LATINO, White Blood Cell	0.0257	3.0000	< 0.0001
Blood Gas, White Blood Cell, DIVORCED	0.0371	2.6000	< 0.0001
Blood Gas, White Blood Cell, lessthan40	0.0857	2.1429	< 0.0001

DIVORCED, Blood Gas, F	0.0314	3.6667	0.0023
Sedimentation Rate, Blood Gas, F	0.0343	4.0000	0.0016

Within listed highly contrast patterns, there are 16 unique items, three Age ranges, two Sex, one Mortality Status, four Marital Status, two Races, and four Lab Tests (Blood Gas, Sedimentation Rate, White Blood Cell, D-Dimer). All four of the discovered lab tests are related to diagnosis of PE. While a single lab test is non-specific, the combination of these lab tests could indicate a high risk of PE. The highest growth rate is 7 which means the pattern occurs 7 times higher in PE group rather than non-PE group. Of the 19 identified contrast patterns, 17 retain statistical significance.

I. Implementation Considerations

A. Human Interaction, Application, and Visualization

The primary goal of this research is to develop a diagnostic algorithm that relies on the continuous ECG Lead II waveform readings that typically form a part of normal care for patients in critical care units. Because we will rely on this standard of care, little extra interaction is necessary for clinicians to take advantage of a potential CDS rule that fires on the basis of these data. Delivery of an alert triggered by the CDS can be made in many ways, depending on the specifics of the electronic medical record (EMR) setup. Direct SMS alerts can be sent to the attending physicians phone, or to a rapid response critical care team. Additionally, alerts can be sent via hospital paging systems, integrated EMR messaging apps on smart phones, or via on screen alerts on critical care dashboards.

Because PE is normally diagnosed with the aid of ECG readings, it would be appropriate to show the progression of the amplitude changes in the waveform components, specifically the QRS segment. A comparison of baseline ECG reading to the reading that triggered the alerts would elucidate the progression to a clinician trained in reading ECG waveforms.

B. Integration

Because of the high resolution of the waveform data (up to 12 Hz), a streaming protocol for sending newly acquired data to a rule processor will be necessary to handle the continuously changing data. The proposed algorithm can take advantage of existing protocols that stream data such as ECG to central critical care unit monitors. Defining an additional destination for the data will allow the algorithm to receive real-time ECG data for processing.

II. Evaluation

C. Results

This is the first retrospective study we know of combining the use of ECG waves from bedside monitors with existing clinical lab test to process early identification of PE in ICU patients on MIMIC III, a heterogeneous, data set. We will make the interpretation and clinical evaluation according to the “Harrison’s Principles of Internal Medicine” textbook, 19th edition, and “Evaluation of Patients with Suspected Acute Pulmonary Embolism: Best Practice Advice from the Clinical Guidelines Committee of the American College of Physicians”. The results are mainly using clinical lab tests items and ECG waveforms which could help diagnose PE. We are going to observe the dynamic changes in both lab tests values and Lead II, QRS complex ECG wave patterns, and detect if the lab tests and ECG waves could show the PE trend. The recorded time for lab tests and ECG waves in the dataset could help us to determine when the PE started to happen. Considering the recorded timestamp of a PE diagnosis may not accurately reflect the actual onset in our dataset, the main metric of interest will be the accuracy with which we can distinguish between PE and non-PE patients.

D. Clinical Findings

Our clinical model analysis shows that there are 19 rules associated with PE. As per the recommendation from Pulmonary Embolism-Third Edition [37], we assume some clinical attributes, like arterial blood gas, white blood cell count (WCC), ESR (erythrocyte sedimentation rate), D-dimer, LDH (Lactic acid dehydrogenase), CPK(Creatine phosphokinase), AST (aspartate aminotransferase), B-type natriuretic peptide (BNP), amino-terminal B-type natriuretic peptide (NT-proBNP) and FDP (Fibrin and fibrinogen degradation product) are highly associated with PE diagnosis. However, there are 12 rules involving abnormal arterial blood gas; 9 rules involving increased ESR; 5 rules containing marital status-Divorced; 4 rules involving increased WCC; 2 rules including age, less than 40yrs; and 1 rule having increased d-dimer (> 500ng/ml). We didn't observe any rule including LDH, CPK, AST, FDP, BNP, or NT-proBNP.

A partial pressure of oxygen(PaO2) blood gas is a measurement of how much oxygen is in your blood. It is also indicative of the acidity (pH) of your blood. PaO2, when low in patients with suspected acute PE, has been shown to be a very helpful adjunct in the diagnostic assessment [38]. Even though the Stein study found that abnormal PaO2 may be included among clinical and laboratory findings in patients with PE, it might be normal in some patients with PE.

As we know, the ESR value is an acute phase reactant; D-dimer is a fibrinogen catabolite. And these tests are widely used in PE diagnosis. In our rules, we observed the ESR has the higher frequency than D-dimer. However, D-dimer had 96.7% sensitivity; 67.9% specificity in PE diagnosis [39]. Then we looked it back from our dataset. We found only 48 of 329 PE patients did D-dimer tests, which are much less than the ESR tests. We also consulted with University Hospital MICU physicians, and they indicated that MICU patients mostly had CT/CTA scan upon admission to ICU, which are gold standards to diagnose PE.

WCC/ Leukocytosis isn’t a specific indicator in acute PE [40]. The frequency of an increased WCC was not higher in patients with the pulmonary hemorrhage/infarction syndrome than in patients with acute PE who did not have. One of the reasons is the prevalence of leukocytosis in PE has been unclear. Although, an extensive number of investigations of PE over a decade may relate to the fact that acute PE is usually associated with other complications, which themselves may cause leukocytosis.

Besides the clinical lab tests, the divorced marital status is also related to PE. A lot of the scientific evidence showed the association between divorce and health. A high-quality social relationship is positively related to increased life satisfaction and psychological well-being [41] and negatively associated with morbidities and mortality from a range of disease processes [42]. The divorced status may cause unhealthy living

habit; like depression, drug or alcohol addiction, high-calorie diet, and less movement, which are high-risk of PE.

In the meanwhile, we didn't observe any rule contain LDH, CPK, AST, FDF, BNP, and NT-proBNP. These clinical attributes are strongly recommended by medical textbook or PE guidelines. We surmise that they didn't show up for the following reasons. Firstly, they are all biomarkers which indicate organ damage or indirect metabolism markers. They may not be sensitive enough to show in our clinical model. In addition, we were using a single-center ICU dataset, and the amount of the PE patients, and consequently incidences of these lab results, is limited. These lab tests might emerge when we are using multi-center dataset.

E. ECG Findings

Presented by most current studies, ECG changes may be indicative of right ventricle/ventricular (RV) strain. Examples include T wave inversion in leads V1-V4; right atrial enlargement (P pulmonale) – peaked P wave in the lead II > 2.5 mm in height; SI QIII TIII pattern – deep S wave in the lead I, Q wave in III, inverted T wave in III. This "classic" finding is neither sensitive nor specific for pulmonary embolism but can be found in 20% of patients with PE. No research has mentioned or recommend using lead II QRS complex as a PE risk predictor. In our ECG model, we had an accuracy of 0.6 by using a neural network. From data mining perspective, it is not considered an ideal result; however from the clinical perspective, it is optimistic. Clinically, the ECG wave of PE is non-specific. There isn’t any other study with strong evidence to recommend any ECG pattern for diagnosis of PE. But in our model, we see some evidence that the trend of the lead II QRS complex could be used to assist the early detection of PE.

Conclusion

Both LSTM neural network models on LEAD III ECG and Contrast Mining have produced results which have identified the potential usefulness of non-traditional methods for confirming a diagnosis of PE. Importantly, the QRS Complex of Lead II ECG waveforms have not traditionally been associated with diagnosis of PE. However, the fact that the Lead II ECG segments were able to show a small gain in accuracy over random chance of classifying PE patients using a small training dataset justifies further research into this area.

References

[1] Office of the Surgeon General, *The Surgeon General's Call to Action to Prevent Deep Vein Thrombosis and Pulmonary Embolism*. 2008.

[2] K. L. Buesing, B. Mullapudi, and K. A. Flowers, "Deep Venous Thrombosis and Venous Thromboembolism Prophylaxis," *Surgical Clinics of North America*, vol. 95, no. 2. Elsevier, pp. 285–300, 01-Apr-2015.

[3] J. Caprini, "Risk assessment as a guide to thrombosis prophylaxis.," *Curr. Opin. Pulm. Med.*, vol. 16, no. 5, pp. 448–452, Sep. 2010.

[4] G. E. Raskob *et al.*, "Thrombosis: a major contributor to the global disease burden," *J. Thromb. Haemost.*, vol. 12, no. 11, pp. 1580–1590, Nov. 2014.

[5] P. D. Stein and F. Matta, "Epidemiology and incidence: The scope of the problem and risk factors for development of venous thromboembolism," *Clinics in Chest Medicine*, vol. 31, no. 4. Elsevier, pp. 611–628, 01-Dec-2010.

[6] R. S. Wiener, L. M. Schwartz, and S. Woloshin, "Time Trends in Pulmonary Embolism in the United States," *Arch. Intern. Med.*, vol. 171, no. 9, pp. 831–837, May 2011.

[7] D. Mozaffarian *et al.*, "Heart disease and stroke statistics-2016 update: A Report from the American Heart Association," *Circulation*, vol. 133, no. 4. pp. e38–e48, 26-Jan-2016.

[8] K. T. Horlander, D. M. Mannino, and K. V. Leeper, "Pulmonary embolism mortality in the United States, 1979-1998: an analysis using multiple-cause mortality data.," *Arch. Intern. Med.*, vol. 163, no. 14, pp. 1711–1717, Jul. 2003.

[9] F. A. Anderson *et al.*, "A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism: The Worcester DVT study," *Arch. Intern. Med.*, vol. 151, no. 5, pp. 933–8, May 1991.

[10] D. M. Courtney and J. A. Kline, "Identification of prearrest clinical factors associated with outpatient fatal pulmonary embolism," *Acad Emerg Med*, vol. 8, no. 12, pp. 1136–1142, Dec. 2001.

[11] J. Fanikos, A. Rao, A. C. Seger, D. Carter, G. Piazza, and S. Z. Goldhaber, "Hospital costs of acute pulmonary embolism," *Am. J. Med.*, vol. 126, no. 2, pp. 127–132, Feb. 2013.

[12] J. C. LaMori, O. Shoheiber, S. H. Mody, and B. K. Bookhart, "Inpatient resource use and cost burden of deep vein thrombosis and pulmonary embolism in the United States," *Clin. Ther.*, vol. 37, no. 1, pp. 62–70, Jan. 2015.

[13] G. H. Guyatt, E. A. Akl, M. Crowther, D. D. Gutterman, and H. J. Schuünemann, "Executive summary: antithrombotic therapy and prevention of thrombosis: American College of Chest Physicians evidence-based clinical practice guidelines," *Chest J.*, vol. 141, no. 2_suppl, p. 7S--47S, 2012.

[14] O. Arseven, C. Sevinç, F. Alataş, N. Ekim, L. Erkan, and S. Fındık, "Türk Toraks Derneği Pulmoner Tromboembolizm Tanı ve Tedavi Uzlaşı Raporu," *Türk Toraks Derg.*, vol. 10, no. 11, pp. 1–49, 2009.

[15] J. Hirsh, G. Guyatt, G. W. Albers, R. Harrington, and H. J. Schunemann, "Antithrombotic and thrombolytic therapy: American College of Chest Physicians evidence-based clinical practice guidelines," *CHEST J.*, vol. 133, no. 6_suppl, p. 110S--112S, 2008.

[16] E. J. R. Van Beek, E. M. J. Brouwers, B. I. N. Song, P. D. Stein, and M. Oudkerk, "Clinical validity of a normal pulmonary angiogram in patients with suspected pulmonary embolism - A critical review," *Clin. Radiol.*, vol. 56, no. 10, pp. 838–842, Oct. 2001.

[17] P. D. Stein *et al.*, "Complications and validity of pulmonary angiography in acute pulmonary embolism," *Circulation*, vol. 85, no. 2, pp. 462–468, Feb. 1992.

[18] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database.," *Sci. data*, vol. 3, p. 160035, May 2016.

[19] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, pp. E215-20, Jun. 2000.

[20] A. Perrier *et al.*, "Non-invasive diagnosis of venous thromboembolism in outpatients," *Lancet*, vol. 353, no. 9148, pp. 190–195, Jan. 1999.

[21] W. Zong, G. B. Moody, and D. Jiang, "A robust open-source algorithm to detect onset and duration of QRS complexes," in *Computers in Cardiology, 2003*, 2003, pp. 737–740.

[22] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities.," *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, Nov. 1982.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Erratum: Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947–951, Jun. 2000.

[25] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[26] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems."

[27] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” in *ACM SIGMOD Record*, 1993, pp. 207–216.

[28] P. K. Novak and N. Lavracˇ, “Supervised Descriptive Rule Discovery: A Unifying Survey of Contrast Set, Emerging Pattern and Subgroup Mining,” *J. Mach. Learn. Res.*, vol. 10, pp. 377–403, 2009.

[29] M. Bayes and M. Price, “An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.,” *Philosophical Transactions (1683-1775)*, vol. 53. Royal Society, pp. 370–418.

[30] P. Domingos and M. Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Los,” *Mach. Learn.*, vol. 29, no. 1, pp. 103–130, 1997.

[31] J. R. Quinlan, “Induction of Decision Trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[32] D. R. Cox, “The Regression Analysis of Binary Sequences,” *J. R. Stat. Soc.*, vol. 20, no. 2, pp. 215–242, 1958.

[33] J. A. Hanley and B. J. McNeil, “A method of comparing the areas under receiver operating characteristic curves derived from the same cases.,” *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.

[34] D. GREEN and J. SWETS, “Signal detection theory and psychophysics.” Los Altos, CA: Peninsula, 1988.

[35] D. M. Hawkins, “The Problem of Overfitting,” *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, 2004.

[36] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in *Appears in the International Joint Conference on Articial Intelligence (IJCAI)*, 1995, pp. 1–7.

[37] P. D. Stein, *Pulmonary embolism*. .

[38] M. M. Szucs, H. L. Brooks, M. D. W. Grossman, and J. E. Dalen, “Diagnostic Sensitivity of Laboratory Findings in Acute Pulmonary Embolism,” *Ann. Intern. Med.*, vol. 74, no. 2, pp. 161–166, Feb. 1971.

[39] N. Steeghs, R. J. Goekoop, R. W. L. M. Niessen, G. J. P. M. Jonkers, H. Dik, and M. V. Huisman, “C-reactive protein and D-dimer with clinical probability score in the exclusion of pulmonary embolism,” *Br. J. Haematol.*, vol. 130, no. 4, pp. 614–619, Aug. 2005.

[40] A. Afzal, H. A. Noor, S. A. Gill, C. Brawner, and P. D. Stein, “Leukocytosis in acute pulmonary embolism.,” *Chest*, vol. 115, no. 5, pp. 1329–1332, May 1999.

[41] E. Diener and M. E. P. Seligman, “Very Happy People,” *Psychol. Sci.*, vol. 13, no. 1, pp. 81–84, Jan. 2002.

[42] B. N. Uchino, “Social Support and Physical Health: Understanding the Health Consequences of Relationships,” vol. 161, no. 3, p. 222, 2004.