Call me by your name

โดย ผศ. นพ. ปิยพงษ์ คำริน, นพ. ณรงค์ศักดิ์ แตงอ่อน, และ น้องเพียว

# Objective

วินิจฉัยแยกภาวะ septic shock ออกจาก septicaemia โดยใช้ machine learning models

# Abstract

# Introduction

Septic shock เป็นภาวะวิกฤต ที่ต้องอาศัยการวินิจฉัยอย่างรวดเร็วเพื่อการรักษาที่เหมาะสม ในปัจจุบัน การวินิจฉัย septic shock จากผู้ป่วยที่มีภาวะ septicaemia ใช้เกณฑ์การวินิจฉัย เช่น SOFA score, SIRS score ในการตัดสินใจ อย่างไรก็ตาม การใช้เกณฑ์การวินิจฉัยดังกล่าวเพียงอย่างเดียว อาจจะไม่ได้เหมาะสมในผู้ป่วยทุกราย เนื่องจากผู้ป่วยแต่ละคนมีปัจจัยพื้นฐาน ความเจ็บป่วย และความรุนแรงของโรคที่แตกต่างกัน

ด้วยข้อจำกัดดังกล่าว จึงเป็นที่มาในการวิจัยในการหาปัจจัย และ ความสัมพันธ์ของปัจจัยดังกล่าว ในการช่วยวินิจฉัยแยกภาวะ septic shock ออกจาก septicaemia

# Materials and Methods

โดยการวินิจฉัยแยกภาวะ septic shock ออกจาก septicaemia นั้นจะใช้ machine learning models ช่วยในการตัดสินใจ โดยทำการฝึก machine learning models ให้เรียนรู้จากข้อมูลผู้ป่วย ที่ได้ระบุ septic shock และ septicaemia เรียบร้อยแล้ว เพื่อสร้างองค์ความรู้ และ แนวทางการวินิจฉัยแยกโรคโดยอัตโนมัติ โดยขั้นตอนดังกล่าวแบ่งเป็น 4 ลำดับดังนี้

1. ดึงข้อมูลผู้ป่วยที่ได้รับการวินิจฉัย septic shock และ septicaemia จากบันทึกสรุปเวชระเบียน
2. คัดแยกคำสำคัญออกจากบันทึกสรุปเวชระเบียน
3. เลือกเฉพาะปัจจัยที่สำคัญที่ช่วยในการวินิจฉัยแยกโรค
4. ฝึก machine learning models ให้เรียนรู้แนวทางการวินิจฉัยแยกโรค

### ดึงข้อมูลผู้ป่วยที่ได้รับการวินิจฉัย septic shock และ septicaemia จากบันทึกสรุปเวชระเบียน

ตัวอย่างการทดลอง ได้จาก บันทึกสรุปเวชระเบียน (discharge summary) จำนวน 4,489 ตัวอย่าง จากฐานข้อมูลอิเล็กทรอนิกส์ ของ โรงพยาบาลมหาราชนครเชียงใหม่ แบ่งเป็น ผู้ป่วยที่มีภาวะ septicaemia ร่วมกัน septic shock 1,963 คน และ ผู้ป่วยที่มีภาวะ septicaemia อย่างเดียว 2,526 คน ในปี 2560

*“Female 78 years old &lt;br/&gt;\*\*\* present with Shock from UTI &lt;br/&gt; 2 days leg edema receive diuretic and muscle relaxant &lt;br/&gt; 1 days feel epigastric discomfort while watching TV &lt;br/&gt; went to community hospital EKG : AF rate 46 /min repeat EKG : ST-T change treat with &lt;br/&gt; Enoxaparin 12 mg sc and ASA gr V Follow up EKG : new RBBb refer onLevophed + adrenaline&lt;br/&gt; Septic work up : leukocytosis UA 50-100…”*

ตัวอย่างบันทึกสรุปเวชระเบียน

### คัดแยกคำสำคัญออกจากบันทึกสรุปเวชระเบียน

1. Tokenization

เป็นการตัดคำจากประโยคโดยใช้ช่องว่างระหว่างคำ โดยแต่ละคำจะเรียกว่า feature ตัวอย่างเช่นในตัวอย่างบันทึกสรุปเวชระเบียนจะได้ชุดคำดังนี้

[Female, 78, years, old, &lt;br/&gt;\*\*\*,….]

1. Lower case

จากนั้นเปลี่ยนตัวอักษรทุกตัวเป็นตัวพิมพ์เล็ก

[female, 78, years, old, &lt;br/&gt;\*\*\*,….]

1. SOFA and SIRS criteria extraction

ใช้ regular expression ในการดึงข้อมูล blood pressure, respiratory rate, temperature, PaO2, platelet, white blood cell count, serum creatinine (ไม่แสดงรายละเอียดในที่นี้)

1. Remove stop words

ตัดคำเชื่อมในภาษาอังกฤษ ที่ไม่มีความสำคัญต่อการวินิจฉัยแยกโรคออก เช่น a, an, the, and, are, is เป็นต้น

1. Lemmatization and Stemming

Lemmatization คือการเปลี่ยนคำให้เป็นคำที่เป็นรากศัพท์ (สนใจ grammar) เช่น studies เป็น study

Stemming คือการตัด prefix หรือ suffix ของคำออก ให้เหลือแต่คำตั้งต้น (ไม่สนใจ grammar) เช่น studies เป็น studi

1. Remove non-relevant features

เรียงลำดับ feature จากค่า Pearson correlation ระหว่าง feature กับ septic shock และ septicaemia แล้วเลือก 100 feature แรก

จากนั้น พิจารณาเลือกบาง feature ที่ไม่มีความหมายทางคลินิก ใน 100 ตัวแรกออก

1. Target class

หมายถึง label ที่บอกว่าเคสนั้นเป็นอะไร ในที่นี้มีสองแบบ คือ

1. septicaemia with shock หมายถึง มีคำว่า septicaemia และ shock ปรากฏใน 1 ใน 3 ของโรคที่วินิจฉัยในผู้ป่วยรายนั้น

2. septicaemia without shock หมายถึง มีคำว่า septicaemia แต่ไม่มี shock ปรากฏใน 1 ใน 3 ของโรคที่วินิจฉัยในผู้ป่วยรายนั้น

# Result

จำนวนผู้ป่วยที่ใช้ในการทดลอง

Feature (ปัจจัย หรือ ข้อมูลที่ใช้ในการพิจารณาการวินิจฉัยแยกโรค)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Features | No | Features | No | Features | No | Features | No | Features |
| 1 | septic | 21 | regular | 41 | secretion | 61 | g/s | 82 | subicu |
| 2 | bp | 22 | rt | 42 | anc | 62 | muddy | 83 | airway |
| 3 | er | 23 | hemodynamic | 43 | jx | 63 | lavage | 84 | echo |
| 4 | load | 24 | pale | 44 | cast | 64 | na | 85 | start |
| 5 | line | 25 | metabolic | 45 | cpr | 65 | h/d | 86 | lt. |
| 6 | nss | 26 | ward | 46 | bma | 66 | baseline | 87 | ceftazidime |
| 7 | levophed | 27 | pr | 47 | sound | 67 | ft3 | 88 | cloxacillin |
| 8 | drop | 28 | empirical | 48 | ua | 68 | ugih | 89 | tft |
| 9 | central | 29 | atn | 49 | blood | 69 | consult | 90 | sub |
| 10 | heart | 30 | drip | 50 | sputum | 70 | set | 91 | hco3 |
| 11 | neutropenia | 31 | soft | 51 | serum | 71 | lpm |  |  |
| 12 | imp | 32 | lungs | 52 | yr | 72 | wbc= |  |  |
| 13 | aki | 33 | rr | 53 | =-ve | 73 | chemo |  |  |
| 14 | febrile | 34 | bur/cr | 54 | sls2 | 74 | tracheostomy |  |  |
| 15 | abd | 35 | hypovolemic | 55 | gross | 75 | cirrhosis |  |  |
| 16 | cvp | 36 | access | 56 | creatinine | 76 | aml |  |  |
| 17 | hydrocortisone | 37 | edema | 57 | stable | 77 | ft4 |  |  |
| 18 | murmur | 38 | crrt | 58 | med | 78 | male |  |  |
| 19 | nephron | 39 | u/d | 59 | sec | 79 | advice |  |  |
| 20 | acidosis | 40 | meropenem | 60 | fine | 80 | cardiogenic |  |  |

### ตัวอย่างของเคสที่ใช้ในการฝึก machine learning models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | septic | bp | er | load | line | nss | levophed | …………………………… | Target class |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | …………………………………. | no\_shock |
| 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | …………………………………. | shock |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | …………………………………. | no\_shock |
| 4 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | …………………………………. | no\_shock |

0 = ไม่มีคำนี้ปรากฏในสรุปเวชระเบียนในเคสนั้นม 0 = มีคำนี้ปรากฎในสรุปเวชระเบียนในเคสนั้น

### Prediction results of machine Learning models

|  |  |
| --- | --- |
| Machine learning models | Overall accuracy |
| JRip rules | 79% |
| J48 decision tree | 78% |
| Logistic | 80% |

ในการทดลองนี้เราจะแสดงเฉพาะ Machine learning model ที่แสดงวิธีการตัดสินแยกโรค เพื่อให้ง่ายต่อการเข้าใจกระบวนการการตัดสินใจ และ แสดง เฉพาะ overall accuracy เพื่อแสดงประสิทธิภาพโดยภาพรวมในการวินิจฉัยแยกโรค

Overall accuracy = (TP + TN) / (TP + TN + FP + FN)

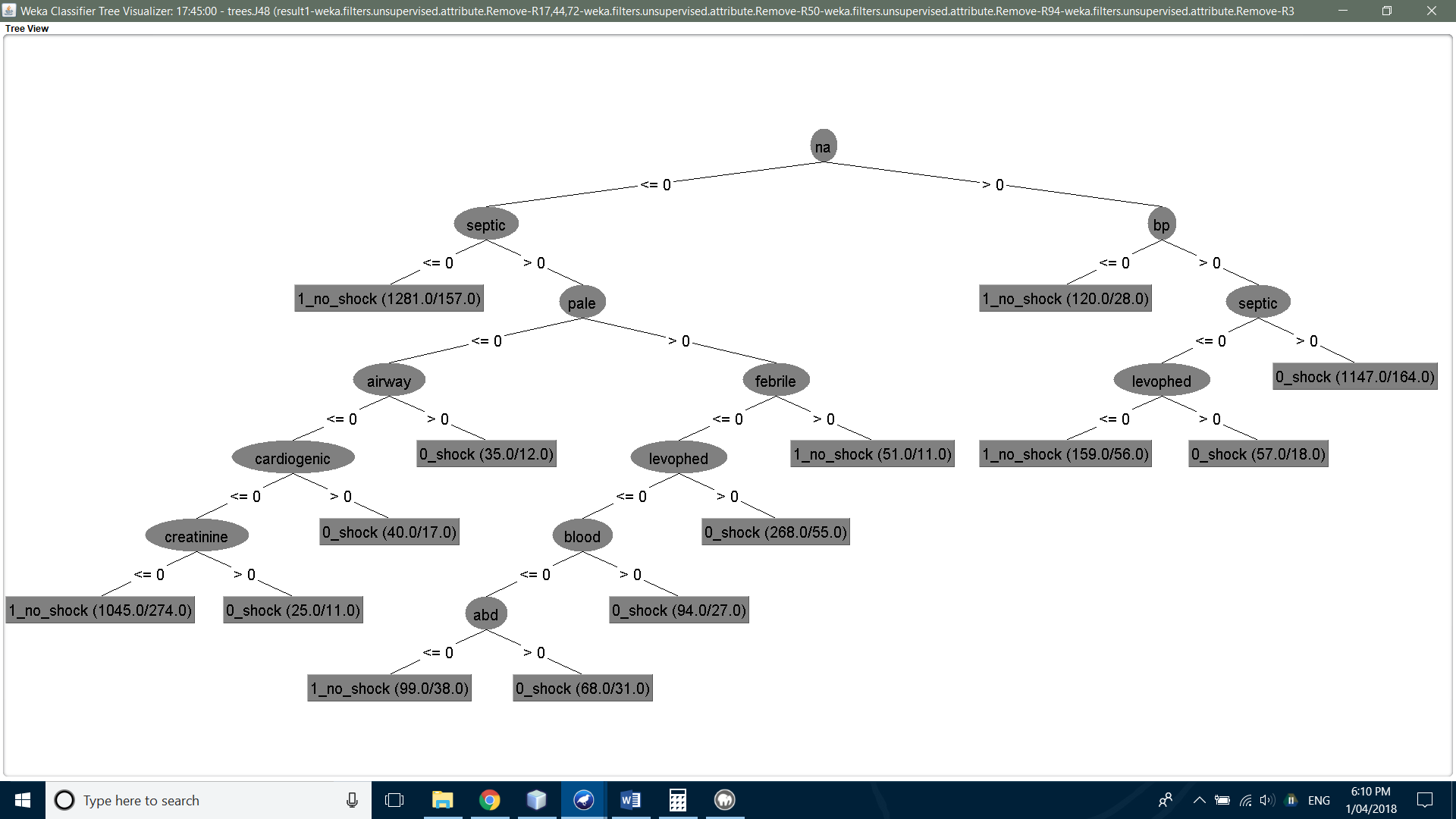
## JRip rules

JRip rules จะสร้างกฎ เพื่อจะใช้ในการตัดสินใจ (คล้ายกับ guideline) โดยกฎที่ JRip สร้าง ออกมาดังนี้

|  |
| --- |
| (na >= 1) and (levophed >= 1) => target\_class=shock (848.0/104.0)  (pale >= 1) and (septic >= 1) and (hco3 >= 1) and (plt\_value >= 10) and (ward >= 1) => target\_class=shock (43.0/0.0)  (heart >= 1) and (septic >= 1) and (na >= 1) => target\_class=shock (288.0/56.0)  (pale >= 1) and (septic >= 1) and (levophed >= 1) => target\_class=shock (278.0/67.0)  (septic >= 1) and (heart >= 1) and (abd >= 1) and (rr\_value >= 20) => target\_class=shock (48.0/9.0)  (bp >= 1) and (pale >= 1) and (septic >= 1) and (male >= 1) => target\_class=shock (60.0/21.0)  (bp >= 1) and (nss >= 1) and (ward >= 1) and (wbc\_value >= 3) and (blood >= 1) => target\_class=shock (18.0/2.0)  (bp >= 1) and (er >= 1) and (med >= 1) and (meropenem >= 1) => target\_class=shock (58.0/23.0)  (septic >= 1) and (heart >= 1) and (blood >= 1) and (er >= 1) and (ml >= 1) => target\_class=shock (21.0/1.0)  (bp >= 1) and (septic >= 1) and (cardiogenic >= 1) => target\_class=shock (39.0/15.0)  (septic >= 1) and (pr >= 1) and (cr\_value >= 2) => target\_class=shock (29.0/8.0)  => target\_class=no\_shock (2759.0/539.0) |

## J48 decision tree

J48 decision tree สร้างแผนผังต้นไม้ ในการวินิจฉัยแยกโรค



ตัวอย่างการแปลผล

1\_no\_shock(1281.0/157.0) = บันทึกของเคสที่ไม่ปรากฏคำว่า “na” และ “sepsis ถูกวินิจฉัยว่าเป็น septicaemia without shock ทั้งสิ้น 1,438 เคส (ถูกต้อง 1281 เคส ไม่ถูกต้อง 157 เคส)

## logistic

ใช้ logistic regression เพื่อทำนาย septic shock ซึ่งจะให้ค่า Coefficients และ Odd ratios จากการวิเคราะห์ดังนี้

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Coefficients** | **Odd ratios** | Features | **Coefficients** | **Odd ratios** | Features | **Coefficients** | **Odd ratios** | Features | **Coefficients** | **Odd ratios** | Features | **Coefficients** | **Odd ratios** |
| **abd** | 0.45 | 1.5684 | cloxacillin | -1.2833 | 0.2771 | heart | 0.4825 | 1.6202 | murmur | 0.3487 | 1.4172 | stable | 0.0982 | 1.1032 |
| **access** | -0.2972 | 0.7429 | consult | -0.2711 | 0.7626 | hemodynamic | 0.115 | 1.1219 | na | 0.6298 | 1.8773 | start | -0.2407 | 0.7861 |
| **acidosis** | -0.0626 | 0.9393 | cpr | 0.1613 | 1.1751 | hydrocortisone | 0.7318 | 2.0789 | nephro | -0.0193 | 0.9808 | sub | 0.3347 | 1.3975 |
| **advice** | 0.0276 | 1.0279 | creatinine | 0.0779 | 1.081 | hypovolemic | 0.4196 | 1.5213 | neutropenia | -0.425 | 0.6537 | subicu | -0.5404 | 0.5825 |
| **airway** | 0.4917 | 1.6351 | crrt | 0.4123 | 1.5103 | iii | -0.2351 | 0.7905 | nss | 0.5743 | 1.776 | tft | -0.3689 | 0.6915 |
| **aki** | 0.2891 | 1.3352 | cvp | -0.3856 | 0.68 | imp | 0.0453 | 1.0464 | pale | 0.6236 | 1.8657 | tracheostomy | -0.233 | 0.7921 |
| **aml** | 0.1738 | 1.1898 | drip | 0.1994 | 1.2207 | iv | -0.1482 | 0.8622 | pr | 0.1287 | 1.1373 | u/d | 0.261 | 1.2982 |
| **anc** | 0.0846 | 1.0882 | drop | 0.0536 | 1.0551 | jx | -0.1275 | 0.8803 | regular | -0.0737 | 0.929 | ua | -0.2531 | 0.7764 |
| **atn** | -0.2108 | 0.81 | echo | 0.3584 | 1.4311 | lavage | 0.2063 | 1.2292 | rr | 0.2894 | 1.3357 | ugih | -0.437 | 0.646 |
| **baseline** | 0.1777 | 1.1944 | edema | 0.0515 | 1.0529 | levophed | 0.5961 | 1.8151 | rt | 0.2978 | 1.347 | ward | 0.0149 | 1.015 |
| **blood** | 0.4472 | 1.5639 | empirical | -0.1507 | 0.8601 | line | 0.1075 | 1.1135 | s1s2 | 0.2797 | 1.3227 | wbc= | -2.1122 | 0.121 |
| **bma** | -0.2718 | 0.762 | er | 0.2387 | 1.2696 | load | -0.0068 | 0.9932 | sec | 0.1885 | 1.2075 | bp\_value | 0 | 1 |
| **bp** | 0.1619 | 1.1758 | febrile | -0.4414 | 0.6431 | lpm | 0.0362 | 1.0368 | secretion | 0.207 | 1.23 | rr\_value | 0.0008 | 1.0008 |
| **bun/cr** | -0.1266 | 0.8811 | fine | -0.3496 | 0.705 | lt. | 0.1373 | 1.1471 | segment | -0.0123 | 0.9877 | temp\_value | -0.0016 | 0.9984 |
| **cardiogenic** | 1.2147 | 3.3693 | ft3 | 1.8691 | 6.4825 | lungs | -0.1819 | 0.8337 | septic | 1.4295 | 4.1766 | pao2\_value | 0.0022 | 1.0022 |
| **cast** | -0.3324 | 0.7172 | ft4 | -0.7168 | 0.4883 | male | 0.3784 | 1.4599 | serum | 0.163 | 1.177 | plt\_value | 0 | 1 |
| **ceftazidime** | -0.262 | 0.7695 | g/s | 0.5253 | 1.691 | med | 0.364 | 1.4391 | set | -0.1322 | 0.8762 | wbc\_value | 0 | 1 |
| **central** | -0.3053 | 0.7369 | gross | 0.2914 | 1.3383 | meropenem | -0.1592 | 0.8528 | soft | -0.1283 | 0.8796 | cr\_value | 0.001 | 1.001 |
| **chemo** | -0.739 | 0.4776 | h/d | -0.4797 | 0.619 | metabolic | -0.1891 | 0.8277 | sound | -0.0611 | 0.9407 |  |  |  |
| **cirrhosis** | 0.0279 | 1.0283 | hco3 | 0.3745 | 1.4543 | muddy | -0.0722 | 0.9303 | sputum | -0.1053 | 0.9001 |  |  |  |

จากกฎการตัดสินใจที่ได้จาก machine learning models ทั้งสามโมเดล สามารถสรุปได้ว่า เมื่อมีคำสำคัญใดปรากฏในบันทึกสรุปเวชระเบียนแล้ว ผู้ป่วยรายนั้นจะมีแนวโน้มที่จะเป็น หรือไม่เป็น septic shock ในอัตราความถูกต้องประมาณ ร้อยละ 80

## JRip rules

|  |
| --- |
| ผู้ป่วยมีภาวะ septic shock เมื่อ ... |
| มี lab serum sodium, platelet, HCO3 (จาก arterial blood gas), WBC |
| มี order levophed, meropenem, nss |
| เพศชาย: male  มีคำเกี่ยวกับ CVS: pale, cardiogenic, heart, pr, bp, blood  Respiratory rate >= 20  Cr >= 2  คำอื่นๆ: septic, ward, med, er, abd |

## J48 decision tree

|  |
| --- |
| ผู้ป่วยมีภาวะ septic shock เมื่อ ... |
| มี serum Na, bp, septic, levophed, blood, abd, airway, cardiogenic, creatinine |

## logistic

|  |
| --- |
| ผู้ป่วยมีภาวะ septic shock เมื่อ ... |
| 1. มีคำว่า abd, advice, airway, aki, aml, anc, baseline, blood, bp, cardiogenic, cirrhosis, cpr, creatinine, crrt, drip, drop, echo, edema, er, ft3, g/s, gross, hco3, heart, hemodynamic, hydrocortisone, hypovolemic, imp, lavage, levophed, line, lpm, lt., male, med, murmur, na, nss, pale, pr, rr, rt, s1s2, sec, secretion, septic, serum, stable, sub, u/d, ward ปรากฎในบันทึก 2. High RR, PaO2, Cr 3. Low Temp |

# Discussion

# Conclusion