Identification of Disease States for Trauma Patients using Commonly Available Hospital Data

Mahnaz Koupaee  
Department of Computer Science,  
University of California, Santa Barbara  
Santa Barbara, CA 93106  
koupaee@cs.ucsb.edu

Yuanyang Zhang  
Department of Computer Science,  
University of California, Santa Barbara  
Santa Barbara, CA 93106  
yuanyang@cs.ucsb.edu

Mitchell Cohen  
Department of Surgery,  
Denver Health and Hospital Authority,  
777 Bannock St, Denver, CO, 80204, USA.  
mitchell.cohen@dhha.org

Linda Petzold  
Department of Computer Science,  
University of California, Santa Barbara  
Santa Barbara, CA 93106  
petzold@cs.ucsb.edu

Tie Bo Wu  
Department of Mechanical Engineering,  
University of California, Santa Barbara  
Santa Barbara, CA 93106  
tiebo@ucsb.edu

Abstract—Trauma is one of the main causes of hospitalization. Time is of the essence in diagnosis and treatment of trauma patients with severe injuries. To assist in decision-making, we propose a hidden Markov model for identification of disease states through which patients progress. An important property of our model is that it is based on features which are routinely collected in hospital trauma centers. Using a hidden Markov model based on fifteen features, six different patient states are identified. The resulting Markov model can be useful in identifying patients’ states to assist in diagnosis and treatment.

Keywords—hidden Markov model, trauma, mortality, MIMIC dataset

I. INTRODUCTION

Trauma is the leading cause of death between the ages of 1- and 54 and has he highest morbidity for those who survive [1]. Different causes and severities of trauma have their own characteristics of injury, physiology and biology requiring different approaches to resuscitation and treatment. Categorizing trauma patients into physiologic states and knowing the possible transitions between these states would be of huge importance providing actionable clinical decision support. In this paper, we infer the states from patients’ time-series data. Among these states, we are particularly interested in coagulopathic states because of their high mortality. Coagulopathy is a condition in which blood fails to clot properly, therefore blood loss is accelerated. Being able to understand the progression of patient states using a hidden Markov model can assist medical care providers by alerting them of the need for interventions when the patient’s state changes. Earlier work of Zhang et al. [2] has addressed this problem; however, the hidden Markov model was inferred from only blood-related factors. These tests are usually more costly than other measurements such as the ones used in this study, and are not commonly performed in a clinical setting. Moreover, we are interested in knowing how well widely available features can represent the physiologic states of trauma patients.

In this study, we have used the latest version of a publicly available dataset called MIMIC [3]. MIMIC comprises de-identified health data associated with more than 40,000 critical care patients. We have selected fifteen features that are measured at different time points, and applied a hidden Markov model that generated six different states.

The contributions of this study are twofold. First, by applying a hidden Markov model we infer the different states that trauma patients transition through on their way to recovery or death. Having the states and transitions, and understanding the progression through states can enhance the decision-making process. Second, we identify states in which early intervention may be critical to influencing patient outcomes. This paper is organized as follows. The dataset and preprocessing used in this work are described in Section 2. Section 3 describes the methods briefly. Section 4 discusses the results generated by our model. In Section 5 we provide some remarks of the work and finally, Section 6 concludes the paper.

II. DATASET AND METHODS

We made use of data in the MIMIC-III dataset v1.3 [3], which is a freely available dataset containing medical records for Intensive Care Unit (ICU) patients at the Beth Israel Deaconess Medical Center between 2001 and 2012. It includes demographics, vital signs, laboratory tests, medications, and more. This dataset also includes notes describing patients’ states during different times.

A. Preprocessing and data extraction

As we are interested in trauma patients, we extracted, time-series measurements (different factors measured at different time points such as heart rate, blood pressure, etc.) and mortality (whether a patient died or survived) for every trauma patient, using ICD-9 codes [4]. To retrieve the data of trauma patients, we used the ICD-9 codes ranging from 800 to 999, which are related to trauma. We aggregated the measurements for each patient over 4-hour time slots. We wanted to make sure that there are enough measurements taken for a patient so that the model can make use of the provided information for its predictions. Using shorter intervals would create more missing values, while using longer intervals might ignore some important measurements.

The resulting trauma data includes about 70,000 records with more than 900 features. Due to data sparsity, several additional steps of preprocessing were performed. Columns representing features with less than 2,000 non-missing values were dropped so that the ones measured for a large number of patients remained. Moreover, features were processed to remove duplicates. Later, correlation analysis (described in the Supplement A) of the remaining features and selection of only one of the correlated features resulted in fifteen features.
Table 1 in the Supplement B shows the feature names and their representations used throughout the paper.

As the hidden Markov model requires the initial hour measurements, patients with no initial measurements were removed. After applying multiple levels of preprocessing, the dataset included 1,275 patients with 6,127 records.

B. Hidden Markov Model

The HMM is widely used in natural language processing [5], speech recognition [6] and for biological sequence analysis [7]. It can be viewed as a specific instance of the state space model in which the latent variables are discrete and can take on K distinct values. The latent variables are the discrete multinomial variables \( z_n \) describing which component of the mixture is responsible for generating the corresponding observation \( x_n \). The probability distribution of \( z_n \) depends on the state of the previous latent variable \( z_{n-1} \) through a conditional distribution \( p(z_n | z_{n-1}) \). The hidden states of the Markov model satisfy the Markov property, i.e., the probability to be in the current state at time \( n \) depends only on the previous state at time \( n-1 \).

In our dataset, observations (for each patient) at time \( t_0 \) are denoted by \( x_{n0} \) vectors containing 15 features representing the different measurements of each patient. The joint probability distribution for each patient over both latent and observed variables is then given by:

\[
p(X, Z | \theta) = p(x_1 | \theta) \prod_{n=2}^{N} p(z_n | z_{n-1}, A) \prod_{n=1}^{N} p(x_n | z_n, \theta) \quad (1)
\]

where \( X = \{x_1, ..., x_N\}, Z = \{z_1, ..., z_N\} \), and \( \theta = \{\pi, A, \varphi\} \) denotes the set of parameters governing the model. \( N \) is the number of observations for each patient. Here, the time interval between two consecutive observations is 4 hours, therefore each \( x_n \) represents the measurements of different features every four hours. \( A \) is the transition matrix representing probabilities of transition between states, \( \pi \) is the initial probability of states and \( \varphi \) includes parameters (\( \mu_k \) the mean vector and \( \Sigma_k \) the covariance matrix) controlling the emission probabilities distribution.

Calculating the parameters of the hidden Markov model using the likelihood function directly is difficult. We therefore turn to the expectation maximization (EM) algorithm for an efficient framework for maximizing the likelihood function in our hidden Markov model [8].

The EM algorithm can be used to find the parameters of our HMM since it involves latent variables in addition to known parameters and known data observations. To find a maximum likelihood solution we need to take the derivatives of the likelihood function with respect to all the unknown values of both the parameters and the latent variables, and simultaneously solve the resulting equations. The derivations result in equations in which the solution to the parameters requires the values of the latent variables and vice versa. Using the EM algorithm, we pick arbitrary values for one of the two sets of unknowns, use them to estimate the second set, then use these new values to find a better estimate of the first set, and then keep alternating between the two until the resulting values both converge to the fixed points.

We train our hidden Markov model using the EM algorithm with the following steps. We first make an initial selection of the parameters \( \theta \) where \( \theta = \{\pi, A, \varphi\} \). For the Gaussian distributions used in this paper, the parameters \( \mu_k \) were initialized by applying the \( K \)-means algorithm to the data, and \( \Sigma_k \) was initialized to the covariance matrix of the corresponding \( K \)-means cluster. We calculated the parameters needed for the maximization step and evaluated the likelihood function. We used the results to find a revised set of parameters \( \theta \) using the M-step equations. We then continued to alternate between E and M steps until the difference of the values of two consecutive likelihood estimates is lower than 0.01.

C. Missing Values

There are two types of missing data in the dataset. The first type usually occurs due to the critical nature of trauma settings where data can be difficult to measure or record. Moreover, patients might be either too sick to get a measurement at some points or too well to require a certain piece of information. This type of missing data which refers to incomplete measurements for patients can be considered as missing at random (MAR) [9]. The second type of missing data is mainly due to data censoring, which occurs when patients are discharged or die. The first type of missing data (MAR) needs to be taken into consideration, while the second type is not a problem and the hidden Markov model can be applied to it directly.

For the MAR data, not only the latent variables but also the missing data are unobserved. The EM algorithm needs to be modified so that the missing data is also considered. The expectation step (E step) can be separated into two parts: expectation with respect to the latent variables and expectation with respect to the missing values. The first step is similar to the situation where missing data is ignored. To find the expectation with respect to the missing values, we calculate the posterior distribution of the missing data. As our data is assumed to be normally distributed, the posterior probabilities will also be normally distributed and their means and covariances can be calculated using the sweep operator [10]. In the M step, the parameters are maximized using evaluated expectations both from latent variables and missing data.

III. RESULTS

In the next following sections, we discuss different results of applying our approach to infer the hidden states which trauma patients transition through on their way to recovery or death.

A. Number of States

To apply the hidden Markov model on our data to infer hidden states from observations, we need to specify the number of states. To select the best number of states, the Markov model was run 10 times for each possible number of
states from 3 to 8 and the BIC (Bayesian Information Criterion) was calculated as follows [2].

\[ BIC = -2 \cdot \log P(X) + \text{params} \cdot \log(ndata) \]  

(2)

where \( ndata \) is the number of samples and \( \text{params} \) is the number of parameters.

\[ \text{Params} = (K-1) + K^* (K-1) + d \cdot K + \frac{d \cdot (d+1)}{2} \cdot K \]  

(3)

Where \( K \) is the number of states. There are \( K - 1 \) values for the initial probability, \( K^* (K - 1) \) parameters representing parameters for the transition probability matrix, \( d \) is the number of features, which is 15, and \( d \times K \) and \( \frac{d \times (d+1)}{2} \times K \) for the mean and covariance matrix for the Gaussian emission probability.

After trying different numbers of states, the model with 6 states, which has the minimum BIC, was selected. We computed the minimum value for each state and plotted the BIC. The minimum BIC value for different numbers of states is shown in Fig. 1.

B. Trauma Patients’ States

After deciding on the number of states, the hidden Markov model was applied and six states with initial probabilities shown in Table 1 were identified.

Table 1: Initial probabilities of hidden Markov model states

<table>
<thead>
<tr>
<th>State</th>
<th>S0</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial probability</td>
<td>0.653</td>
<td>0.155</td>
<td>0.013</td>
<td>0.023</td>
<td>0.14</td>
<td>0.015</td>
</tr>
</tbody>
</table>

The transition probabilities are shown in Fig. 2. Transition probability is the probability of going from one state to another state. Transition probabilities less than 0.01 are not shown in the figure. Each state is represented by a name which describes the characteristics of the state along with the mortality rate associated with it.

Patients can arrive in any of these states, however state S0 is the most likely. Each state and its transitions is shown with a distinct color. To be able to better describe different states, normalized mean values of features are also plotted in Fig. 3.

We also calculated the mortality rates of different states as follows: the mortality feature is calculated based on the in-hospital hour of death. Mortality is assigned to the final state of each patient. To calculate the mortality rate for each state, the number of patients that died is divided by the total number of patients whose last state before death or discharge is that state. To remove the effect of data censoring, we ignored the patients whose death or discharge occurs 24 hours after the last state is recorded. Table 2 shows the calculated mortality rates for different states.

Table 2: Mortality rates of the states

<table>
<thead>
<tr>
<th>State</th>
<th>S0</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>0.017</td>
<td>0.394</td>
<td>0.049</td>
<td>0.033</td>
<td>0.108</td>
<td>0.231</td>
</tr>
</tbody>
</table>

Finally, we summed up posterior probabilities \( P(z_n | x_n) \) of different states across all patients at each time \( t \), to obtain the total number of patients in each state at time \( t \), shown in Fig. 4.
is from a healthy state to an unhealthy one. The probable
stability) or to S3, another healthy state. There is an
either to itself with high probab
destination of states S1 and S5. Transitions out of S2 are
some issues related to clotting time as this state is the
states because of its low mortality. However, there are still
this state also decreases as time passes.
transition is to S4, which is the shock state. Thus,
least healthy state, it is not highly stable and it can transition
fractures. State S0 has the lowest mortality rate over all the
are associated with this state are representative of multiple
features. State S0 has the lowest mortality rate over all the
states. The reason is probably due to its instability, so that
patients transition to other states very quickly. Patients in this
state never transition to states S1 or S5. States S1 and S5,
which will be described later, are states related to bleeding
and coagulation deficiencies. Thus, if a patient arrives in
state S0, this patient is not at that point at immediate risk of
transitioning to an unhealthy state. The most common
transitions for this state are to either S0 to S2 (p=0.146) or S0
to S3 (p=0.152). Some of the patients may also transition to
state S4, which represents the trauma shock state. Figure 5
shows the initial number of patients in state 0 and how
quickly this state transitions to other states. As time passes,
the number of patients in this state decreases.

State S1 (coagulopathic state): This is the unhealthiest
initial state. It is also highly unstable, so timely interventions
may be critical in this state. The values of the features that
are associated with this state are representative of multiple
fractures. State S0 has the lowest mortality rate over all the
states. The reason is probably due to its instability, so that
patients transition to other states very quickly. Patients in this
state never transition to states S1 or S5. States S1 and S5,
which will be described later, are states related to bleeding
and coagulation deficiencies. Thus, if a patient arrives in
state S0, this patient is not at that point at immediate risk of
transitioning to an unhealthy state. The most common
transitions for this state are to either S0 to S2 (p=0.146) or S0
to S3 (p=0.152). Some of the patients may also transition to
state S4, which represents the trauma shock state. Figure 5
shows the initial number of patients in state 0 and how
quickly this state transitions to other states. As time passes,
the number of patients in this state decreases.

State S2 (transitional state): This is one of the healthier
states because of its low mortality. However, there are still
some issues related to clotting time as this state is the
destination of states S1 and S5. Transitions out of S2 are
either to itself with high probability (which shows its
stability) or to S3, another healthy state. There is an
interesting yet unknown transition from state S2 to S1, which
is from a healthy state to an unhealthy one. The probable
explanation is that the bleeding (corresponding to states S1
and S5) is initially controlled but later it occurs unexpectedly
and the state of the patient’s condition worsens (it transitions
back to state S1). Patients may not yet have enough injury or
shock or physiologic perturbation to yet be in sicker state.
More exploration is required to find the causes of this
transition. One possible solution is to go over the notes
written by medical service providers and find the possible
causes. The length of the red bar in Figure 5 increases as
time passes, which shows that more patients will be in this
state as time passes.

State S3 (relatively healthy state): The healthiest and
most stable state of the Markov model is state S3. Most
features are in good shape. All states can transition to this
state. States S0 and S4 can directly go to this state and states
S1 and S5 can transition to S3 via S2. More patients will be
in this state as time passes. The only transition out of S3 is to
S2.

State S4 (shock state): This state is probably associated
with shock. Shock is a life-threatening condition that occurs
when the body is not getting enough blood flow. Lack of
blood flow means that the cells and organs do not get enough
oxygen and nutrients to function properly. Many organs can
be damaged as a result. Shock requires immediate treatment
and can worsen very rapidly. Low blood pressure, low heart
rate and low blood oxygen in state S4 are all possible
symptoms of shock [11]. There are also different causes for
shock. In the case of trauma patients, heart damage, severe
blood loss and spinal cord injury can all result in shock [12].
Very low blood pressure and heart rate are the discriminatory
features for this state. It is also highly stable. States S0, S1
and S5 may transition to this state, so medical service
providers must make timely decisions to prevent this
transition. The only transition out of this state is to S3.

State S5 (acidotic state): Similar to state S1, this state is
quite unhealthy. Features usually have extreme values. Low
pH, CO2 and blood pressure values in addition to high PT,
and potassium are specific to this state. This state can also
represent shock due to low blood pressure. More
significantly, this state is characterized by acidosis, the
condition in which body fluids contain too much acid. Low
pH indicates the acidity of blood. Acidosis occurs when the
kidneys and lungs cannot keep the body’s pH in balance and
it becomes lower than 7.36 [16]. Some experiments suggest
that low pH during the first 24 hours of trauma seems to be a
good predictor for the development of organ failure [13]. The
most common cause of ultra-high potassium (hyperkalemia)
is related to failing kidneys [14]. Also, any condition in
which there is massive tissue destruction can result in
elevated levels of blood potassium, as the damaged cells
release their potassium [15]. Early intervention can make it
possible to transition to other states, specifically to state S2.

IV. DISCUSSION

By applying the hidden Markov model and finding its
parameters, we can estimate states of patients and make
predictions of possible future states and trajectories for new
patients based on their data. Suppose a patient is in state S5
(acidosis), which has a high mortality rate. This state
transitions into state S1 (unhealthy) or state S2 (healthy).
Data like PTT which goes up if the patient transitions to state
S1 (unhealthy) or goes down if transitioning toward state S2
(healthy) can be monitored. Glucose also increases if

Based on the values of features in different states, the
mortality rates calculated in Table 2 and the change in
number of patients in different states shown in Figure 5, we
can consider states 3 and 2 as the healthiest, and states 1, 5
and 4 as the unhealthiest, respectively. The description of
different states is as follows:

State S0 (Early state): This state is the most probable
initial state. It is also highly unstable, so timely interventions
may be critical in this state. The values of the features that
are associated with this state are representative of multiple
fractures. State S0 has the lowest mortality rate over all the
states. The reason is probably due to its instability, so that
patients transition to other states very quickly. Patients in this
state never transition to states S1 or S5. States S1 and S5,
which will be described later, are states related to bleeding
and coagulation deficiencies. Thus, if a patient arrives in
state S0, this patient is not at that point at immediate risk of
transitioning to an unhealthy state. The most common
transitions for this state are to either S0 to S2 (p=0.146) or S0
to S3 (p=0.152). Some of the patients may also transition to
state S4, which represents the trauma shock state. Figure 5
shows the initial number of patients in state 0 and how
quickly this state transitions to other states. As time passes,
the number of patients in this state decreases.

State S1 (unhealthy state): This is the unhealthiest
state, with the highest mortality rate. The clotting-related
features, prothrombin time (PT) and partial thromboplastin
time (PTT), are both high in this state, indicating coagulation
problems. Moreover, some other features have unique
characteristics. In this state, WBC (white blood cell count)
and platelets are the lowest. On the other hand, glucose and
lactic acid have the highest values. Although state S1 is the
least healthy state, it is not highly stable and it can transition
to state S2 with reasonable probability. The other possible
transition is to S4, which is the shock state. Thus,
interventions need to be performed as early as possible to
increase the chance of survival. The number of patients in
this state also decreases as time passes.

State S2 (relatively healthy state): The healthiest and
most stable state of the Markov model is state S3. Most
features are in good shape. All states can transition to this
state. States S0 and S4 can directly go to this state and states
S1 and S5 can transition to S3 via S2. More patients will be
in this state as time passes. The only transition out of S3 is to
S2.

State S3 (relatively healthy state): The healthiest and
most stable state of the Markov model is state S3. Most
features are in good shape. All states can transition to this
state. States S0 and S4 can directly go to this state and states
S1 and S5 can transition to S3 via S2. More patients will be
in this state as time passes. The only transition out of S3 is to
S2.

State S4 (shock state): This state is probably associated
with shock. Shock is a life-threatening condition that occurs
when the body is not getting enough blood flow. Lack of
blood flow means that the cells and organs do not get enough
oxygen and nutrients to function properly. Many organs can
be damaged as a result. Shock requires immediate treatment
and can worsen very rapidly. Low blood pressure, low heart
rate and low blood oxygen in state S4 are all possible
symptoms of shock [11]. There are also different causes for
shock. In the case of trauma patients, heart damage, severe
blood loss and spinal cord injury can all result in shock [12].
Very low blood pressure and heart rate are the discriminatory
features for this state. It is also highly stable. States S0, S1
and S5 may transition to this state, so medical service
providers must make timely decisions to prevent this
transition. The only transition out of this state is to S3.

State S5 (acidotic state): Similar to state S1, this state is
quite unhealthy. Features usually have extreme values. Low
pH, CO2 and blood pressure values in addition to high PT,
and potassium are specific to this state. This state can also
represent shock due to low blood pressure. More
significantly, this state is characterized by acidosis, the
condition in which body fluids contain too much acid. Low
pH indicates the acidity of blood. Acidosis occurs when the
kidneys and lungs cannot keep the body’s pH in balance and
it becomes lower than 7.36 [16]. Some experiments suggest
that low pH during the first 24 hours of trauma seems to be a
good predictor for the development of organ failure [13]. The
most common cause of ultra-high potassium (hyperkalemia)
is related to failing kidneys [14]. Also, any condition in
which there is massive tissue destruction can result in
elevated levels of blood potassium, as the damaged cells
release their potassium [15]. Early intervention can make it
possible to transition to other states, specifically to state S2.

IV. DISCUSSION

By applying the hidden Markov model and finding its
parameters, we can estimate states of patients and make
predictions of possible future states and trajectories for new
patients based on their data. Suppose a patient is in state S5
(acidosis), which has a high mortality rate. This state
transitions into state S1 (unhealthy) or state S2 (healthy).
Data like PTT which goes up if the patient transitions to state
S1 (unhealthy) or goes down if transitioning toward state S2
(healthy) can be monitored. Glucose also increases if
transitional state S1 and down toward state S2. Lactose increases if transitioning to state S1, and decreases toward state 2. Thus, specific features can be identified to monitor which type of transition a patient is experiencing. Similar reasoning can be applied to other states.

While in each state, the increase or decrease of the values of different features may result in a transition to a new state. Table 3 shows the change of features for all possible state transitions. Not only the change of features but also their level of change can result in different states. For instance, a patient who is in state S5 can transition to all states except S3. Two predictive features of state S5, PT and PTT, must decrease so that the state will transition. However, the amount of change in these features can determine the next state more precisely. Moreover, multiple features need to be considered simultaneously. For example, suppose a patient is in state S1, then a decrease of PT and PTT will result in transition to S0, S2 or S4. Therefore, other features such as respiratory rate, blood pressure and heart rate can help in determining what the new state will be.

There is one important point to remember. From this analysis, we cannot determine what causes the transition. We can only say features are associated with the transition. It is possible that we could temporarily move these features towards these of the desired state but that the patient just reverts to his/her earlier state because of some hidden underlying disease mechanisms. On the other hand, the change in features required for the transition may provide a clue that together with a mechanistic understanding of the system, could identify the cause.

Finally, we evaluated the states by looking at the existing literature and checked the reasonability of our states by consulting a trauma surgeon. Since these states are not well-defined in clinical settings and the existing datasets do not contain necessary information describing the states patients transition through, we were not able to conduct any numerical success and error analysis calculating the number of time our model fails or succeeds. Therefore, it is possible that the predicted model by the state at some points does not reflect the true state of a patient. However, considering the trajectory reaching that state may also help determine the current state of the patient.

Table 3: change of features for different state transitions. The first column represents the current states of the patients. The second column gives the states which the current state can transition to. The remaining columns are the features used in the study. (+) sign represents the increase of a feature and (-) sign represents the decrease of a feature in different transitions. For example, the transition from S0 to S3 occurs when the values of PT and PTT are lowered. The (o) sign means no significant change of values.

<table>
<thead>
<tr>
<th>S0</th>
<th>PT</th>
<th>PTT</th>
<th>Resp</th>
<th>HR</th>
<th>WBC</th>
<th>Hemo</th>
<th>Glu</th>
<th>Ph</th>
<th>Pot</th>
<th>BUN</th>
<th>PTT</th>
<th>Lac</th>
<th>CO2</th>
<th>CO2</th>
<th>O2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V. Conclusion

Using a hidden Markov model, we identified six states associated with trauma patients. Based on the type and severity of the injury, patients might have different trajectories and transition through multiple states while in the hospital. Among the inferred states, two (S1, S5) are related to coagulation deficiency. The coagulation-related factors (PT and PTT) are the highest in these two states, and death often occurs due to massive blood loss. Bleeding is one of the major causes of early death in trauma patients and it needs to be controlled as early as possible. Aside from heavy bleeding in state S5, low PH and CO2, high blood urea nitrogen (BUN) and potassium represent other problems such as acidosis, failure of kidneys and or respiration deficiency. Another critical state is S4 which is related to shock. Shock may also occur as a result of a traumatic injury. Knowing which states may transition to the shock state, it is possible that actions to avoid it can be taken. State S0 is mostly associated with blunt trauma, which usually results in multiple fractures. Finally, effective interventions can help trauma patients to transition to the healthy states S2 or S3.

Moreover, our model is able to determine the states even if only some of the measurements are available. Using our model, medical service providers can quickly assess patients’ current states and predict their future conditions and make optimal decisions.

The selection of features for this study was based on the availability in the dataset and the fact that all chosen features are commonly measured in clinical settings. During preprocessing, the removal of features was mainly due to their sparsity and possible correlation with other features. There is a possibility that these features are not all the ones used in hospital procedures, or a different set of features are considered during different stages of a hospital stay to
monitor patients’ conditions. However, our model can be generalized based on different needs so that we can apply our model to various sets of features or even to other diseases and try to find their progression as well. Moreover, our methodological approach is mainly focused on the results for the application under study and not on the complexity of techniques used in achieving those results. This work is the initial attempt, however more complicated techniques may be used for future studies.

REFERENCES


SUPPLEMENT

A. Correlation Analysis

The features available in the dataset may be correlated, but the hidden Markov model assumes features to be independent so we tried to remove the correlation as much as possible. To do so we used the Pearson correlation coefficient and calculated the correlation of all pairs of features. Later, if the correlation coefficient of features was more than some threshold, (we used 0.6 as the threshold), one of the features was kept and the others not used. After calculating the correlation of the features, there were mainly two groups of features: the ones with high correlation (>0.7) and others with correlation coefficient of less than 0.5. Therefore, selecting a threshold of 0.6 was a safe choice so that only one of the highly-correlated features pairs were removed. To select the best feature out of all correlated features, we used each one separately in the analysis to see how the results are improved.

B. Abbreviations

Table 1: Features and their names used throughout the paper

<table>
<thead>
<tr>
<th>Feature</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prothrombin time</td>
<td>PT</td>
</tr>
<tr>
<td>Partial Thromboplastin Time</td>
<td>PTT</td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>Resp</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>BP</td>
</tr>
<tr>
<td>White blood cell count</td>
<td>WBC</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>Hemo</td>
</tr>
<tr>
<td>Glucose</td>
<td>Glu</td>
</tr>
<tr>
<td>Potential of hydrogen</td>
<td>PH</td>
</tr>
<tr>
<td>Platelet count</td>
<td>Platelet</td>
</tr>
<tr>
<td>Heart rate</td>
<td>Heart Rate</td>
</tr>
<tr>
<td>Blood Urea Nitrogen</td>
<td>BUN</td>
</tr>
<tr>
<td>Potassium</td>
<td>Pot</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>Lac</td>
</tr>
<tr>
<td>Carbon dioxide</td>
<td>CO2</td>
</tr>
<tr>
<td>Oxygen</td>
<td>O2</td>
</tr>
</tbody>
</table>