Machine Learning Approach for Improving Decision Support In ICU

SUBMITTED BY:

Mohib Billah Siddiquee ID: 12201043

Mostofa Jamil Fuad ID: 13301081

Md. Fahim Azmain ID: 14201057

Supervisor:

Dr. Md. Ashraful Alam

Assistant Professor

Department of Computer Science and Engineering, BRAC University

SUBMISSION DATE: 22.07.18

Declaration

We, hereby declare that this thesis based on the results found by ourselves by doing several experiment with codes. Materials of works, charts and algorithms found by other researchers are mentioned by reference. This Thesis, neither in whole or in part, has been previously submitted for any degree.

Signature of Supervisor	**Signature of Author**
---	---------------------------------
Dr. Md. Ashraful Alam	
Assistant Professor	
Department of Computer Science and Engineering	---------------------------------
BRAC University	

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgement

This research has been supervised by Dr. Md. Ashraful Alam, Assistant Professor, Department of Computer Science and Engineering, BRAC University. Our supervisor was very dedicated and supportive throughout the whole research work. We also want to thank the BRAC IT for allocating a dedicated computer for us, our work might have undone without that. Lastly, we are thankful to all the faculty members of BRAC University who have inspired and motivated us throughout the entire undergraduate program.

Abstract

Patients in the intensive care unit (ICU) receive a deep observation for controlling and responding to their rapidly changing physiological conditions. The quality of their care depends on clinical staff combining large amounts of clinical data to understand the severity of their illness. Actually in real time doctors and nurses have to take care of huge amount of data. Sometimes, they cannot focus on all the parameters at the same time. After collecting all the parameters, they take decisions. A lack of early recognition of physiologic decline can play a major role in failure to rescue patients. Early prediction is one of the important tasks in the ICU. In this paper, we propose a machine learning approach to improve decision support in ICU. In the proposed model decision tree will be used to predict the future health condition of patients. The curve of the decision tree of the proposed model will show how severe the patient's condition is. It will also show the health improvements and decrements. In this model there is option for controlling the lifesaving machines of ICU like ventilation machine, blood warmer machine and syringe pump machine. To control the machines this model uses logistic regression algorithm. It will use some independent variables to predict the decision of automatic intervention. Using the proposed model doctors can easily monitor the health of ICU patients. As it predicts risks, doctors can take early preparation for worst situation. Automatic intervention decisions for ICU machines can save lives in critical moments. As a whole, the model is specially designed for coronary care unit of ICU.

1. Introduction

1.1 Motivation

The Intensive Care Unit (ICU) is a unit in the hospital where critically ill patients are cared by specially trained staff. Critical illness or injury acutely impairs one or more vital organ systems such that there is a high probability of imminent or life threatening deterioration in the patient's condition. Care of these patients can take place anywhere in the inpatient hospital setting, although it typically occurs in the ICU. Critical care involves highly complex decision making to assess, manipulate, and support vital system functions, to treat single or multiple vital organ system failure, and/or to prevent further life threatening deterioration of the patient's condition the intensive care unit (ICU) treats acutely ill patients in need of radical, lifesaving treatments. ICUs have evolved from the notion that specialized units used for close monitoring and treatment of patients could improve outcomes; many predecessors of the modern ICU were established in the late 1950s to provide respiratory support during a polio epidemic [1]. The importance of mechanical ventilation was mostly realized in the polio epidemic in Copenhagen in 1952 where the mortality rates reduced from 90% to 40% following its introduction [3]. This gradually led to the concept of Intensive Care Units. Intensive care has emerged as a distinct specialty in the world over the last 3-4 decades. ICUs frequently have a high number of staff compared to other hospital departments, and studies have shown reduced incidence of mortality, lower hospital length of stay, and fewer illness complications [2], [3], corroborating the efficacy of the intensive monitoring approach. However, real world constraints restrict the number of nurses and doctors attending to the patients in the ICU [4]. ICUs cost $81.7 billion in the US, accounting for 13.4% of hospital costs and 4.1% of national health expenditures [5]. Most of the developed countries are spending billions of dollars to improve their ICUs. As a developing country Bangladesh is also concerned about this issue. Bangladeshi government and private dwellers has invested a lot to build new ICUs and for developing previous units. Between 2012 and 2016, the number of hospital beds in the Bangladesh increased by 19.2%, but the number of critical care beds increased by 16.5% with occupancy. The first ICU in Bangladesh was established in 1980 at the National Institute of Cardiovascular Diseases (NICVD). Since then the number of ICUs have grown steadily but mostly in the city of Dhaka which is the capital (Fig 1). A total of 44 ICUs were identified in the country. Among them, 40 ICUs were situated in Dhaka city; remaining 4 ICUs were located in other districts of Bangladesh. Of the 40 ICUs of Dhaka, 7 were CCUs with ventilator support, 36 ICUs (90%) were in private hospitals, rest in government hospitals. Total beds in these study hospitals were 8828 and total number of ICU beds was 424 (4.8%). In 1980, there were only 28 ICU beds in Dhaka city. Since then the number of ICU beds have gradually increased. Of all the hospitals studied, 25% hospitals had ≥ 10% beds, and 27.5% hospitals had 5-9% beds dedicated to ICU. Total number of ventilators were 240 in 40 ICUs (i.e. 56.6% ICU beds had accompanying ventilators). 25% ICUs had a ventilator: bed ratio of 1:1. Among the ICUs, 27(68%) were designated as mixed medical and surgical), 7(18%) were CCUs with ventilator, 5(12%) cardiac surgery, and 1(2%) neurology. 21 (64%) ICUs were run by Anesthesiologists, and 4(12%) ICUs by Critical Care Specialists / Intensivists. Rest of the ICUs had Cardiologist, or Neurologist as In-charge. 6 (15%) ICUs were closed ICUs and 34 (85%) were open units. 9 facilities (27%) had on duty doctor: patient ratio of 1: 4. In 8 ICUs (24%), on duty doctor: patient ratio was 1:5. Only 4 ICU (12%) had ratio of 1:3. A nurse: bed ratio of 1:1 was seen in 15 (42%) units. 51% ICU doctors at 27 ICUs and 36% ICU nurses at 32 ICUs were cardio pulmonary resuscitation (CPR) trained. The remaining ICUs failed to furnish the information regarding CPR training of their duty doctors and nurses. Arterial blood gas (ABG) analysis machines were available in 70% units, routine lab facilities in 95% units, bedside echocardiography in 55% units and bedside ultra-solography in 65% units. Among supporting facilities, 19 hospitals had high dependency unit (HDU) support, 10 hospitals had dialysis units and 3 hospitals had CRRT facilities and only one hospital had bed side routine hemodialysis facility. Population of Dhaka City Corporation is 15333571.4. So, there was one ICU bed for approx. 12579 residents of city of Dhaka. Bangladesh must have to develop ICU units to reduce mortality rate and to ensure the health benefit. Unfortunately, in our country there is very little number of research work is happening in this important sector of medical science. On the other hand developed countries are doing a lot of research work in ICU related issues. Moreover, Science is getting advance day by day and people are also finding new diseases. Some of this diseases push people into a hard line between live and death. The ubiquitous monitoring of ICU patients has generated a wealth of data which presents many opportunities but also great challenges. In principle, the majority of the information required to optimally diagnose, treat and discharge a patient is present in modern ICU databases. This information is present in a plethora of formats including lab results, clinical observations, imaging scans, free text notes, genome sequences, hear beat rate continuous waveforms and more. The acquisition, analysis, interpretation, and presentation of this data in a clinically relevant and usable format is the premier challenge of data analysis in critical care [6]. Researchers and scientists are trying to invent new machines and sensors for human body. Some research group is also working to implement Artificial Intelligence and machine learning in this field. There is a possibility of controlling the critical ICU machines from a central unit to reduce the human made error. In our thesis, we highlight how machine learning has been used to address these challenges. We have tried to solve some issues. Basically we tried to control some machines of ICU unit by our own made system.

In ICU there are always critical patients. Some patients are even in life support. They cannot express their emotions and what are they feeling. Doctors and nurses have to take care of them very carefully. Conditions of some ICU's patients change very rapidly. Sometimes patients have more than one issue to monitor. For example, a patient in post heart surgery condition has issue with heart beat rate, blood sugar, blood pressure and many more. Sometimes doctors and nurses find it difficult to merge all these things together in such a critical condition as they are also human. So, we planned this thesis project. This system will be installed in the main computer of an ICU where there will be previous data. For predicting we need a huge dataset. The system will take real time digital data inputs from ICU machineries. There will be a ready moving average data from the data set always. If real time data goes up or down from certain threshold it will start to analyze previous data and depending on that outcome it will use some of machine learning algorithm (decision tree, linear regression) which will predict what is going to happen in next few moments. Then it will give signals or take necessary steps. Firstly, it will control the ventilation system. It means it will give idea to the ventilation machine about when it should be working by itself. Secondly, it will take care of Blood warmer machine. Thirdly, it will control the feeding tube and syringe pump channels. Lastly, we have a plan in future to control the life support system with it for heart patients.

1.2 Literature review

Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques and utilizes factual strategies to give computers the ability to "learn" with data, without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn and utilize for themselves.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict and foresee future events. Beginning from the analysis of a known preparing dataset, the learning calculation creates a inferred function to make forecasts about the output values.

MIMIC-III (‘Medical Information Mart for Intensive Care’) is an extensive, single-focus database containing data identifying with patients admitted to basic care units at a vast tertiary care hospital.

Decision tree is a type of supervised learning algorithm (having a pre-characterized target variable) that is generally utilized as a part of characterization issues. It works not only for categorical but also for continuous input and output variables.

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression's main focus is predictive analysis. Logistic regression is used to portray data information and to clarify the relation between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Patient monitoring refers to the continuous estimation of patient parameters such as heart rate and rhythm, respiratory rate, blood pressure, blood-oxygen saturation, and many other parameters have turned into a typical component of the care of critically ill patients. Whenever precise and immediate decision-making is crucial for effective patient care, electronic monitors frequently are used to collect and display physiological data. Increasingly, such data are collected using non-invasive sensors from less seriously ill patients in a hospital's medical-surgical units, labor and delivery suites, nursing homes, or patients' own homes to detect unexpected life-threatening conditions or to record routine but required data efficiently.

Ventilator or "Breathing Machine" is a machine that does some or all of the major work of breathing for a patient. Ventilators are used to support breathing or aid the functional capacity of the lungs. The machine forces air into the lungs through a tube embedded into the windpipe (trachea) or utilizing a firmly fitting mask. This breathing support might be utilized to help get enough oxygen to the body and to evacuate carbon dioxide. Every Once in a while ICU patients need support from a Ventilator machine as they are not sufficiently wakeful or sufficiently strong to breath safely on their own.

Blood warmers are frequently utilized in different ways as a part of various routes in a critical care setting. Despite the fact that there are diverse makers of such units, all of the gadgets work similarly as the blood tubing is placed in the warmer and the circulation is warmed. The greater part of time the units are utilized as a part of fundamentally sick patients in blend with a rapid infuser machine. The task of such machine is to put the blood into the patient rapidly. Every unit does things a bit in an different way. For instance, there are with the end goal that has arrangements in which the blood is warmed, even in stable patients. Such blood warmers are exceptionally useful in averting damage casualties and therapeutic patients from encountering hypothermia, which is a perilous condition. The machines are expendable, lightweight and require no restrictive tubing. Every one of these highlights makes them simple to operate and transport.

Feeding tubes are a way of giving nutritional sustenance and liquids when a critically ill patient is not able to eat on his own. As a rule, the tube is embedded into the stomach through the nose (nasogastric or NG tube). In some cases feeding tubes are put straightforwardly into the stomach through the skin (gastronomy tube or PEG). The ICU dieticians help to choose the liquid food mixture that will provide to best nutrition to each critically ill patient. This fluid sustenance blend is pumped through the feeding tube into the patient's stomach using a programmable pump.

Syringe pump is particularly intended for those patients that require consistent flow rates and exact injection conveyance. It can be utilized for single or chronic, repeated microinjections through a cannula framework. It has built-in standard syringe parameters, for making the use more convenient. We need to give the name of the infusion and the amount in it through its LCD touch screen and it is programmable.

1.3 Thesis Orientation

So far we have described the motivation that helped us a lot to start our thesis in this topic and after that there is literature review. The rest of this paper is organized as follows: Section 2 describes our methodology and used algorithms. This section also contains description about our used algorithm. We modified some algorithms for our task. There is a clear description about the modification made by us and how it works. We have also described about our datasets. Section 3 has clear description of decision tree and how it works. Then in section 4 there is description of logistic regression and how it involves in our project. In section 5 there is result analysis of our work and the problems we faced in our work. We also explained how we overcame from problems. Then in section 6 we have described our limitations and future plan with this thesis topic. Next in section 7 is for conclusion of the paper. In this section we will discuss about the overall task in brief. At the end we have added all the research paper's and books name we used as reference in our project. As it was not an easy thing to do and it needed too much primary knowledge, we had to take data and information from many reference.

2. Methodology

Main purpose of our thesis is to make some devices of ICU automated. That means it can take its decision without any human intervention. As ICU is very critical area of medical science, we kept in mind always that we have to be very sharp and focused in order to reach our goals. After starting our thesis research task, we visited two hospitals and talked with three ICU specialists. There we tried to figure out the main parameters for a patient in ICU. We also searched internet and talked with some medicine specialist. One thing was very clear that we needed a dataset which contains our necessary parameters. It was not possible to get data directly from ICU. Then we decided to get it from an authentic source. Our algorithms are machine learning algorithm. In any kind of machine learning algorithm database or dataset is an integral part. In this thesis we are using the MIMIC-III critical care database. We could not use this database directly. We collected data from there and made a dataset for our purpose. We planned to use supervised learning because we know what will be the outcome in some cases. Outcome will be chosen between two. Supervised learning was best option for us for automation of ICU machines. Supervised learning is a large area of machine learning that involves learning a mapping between data and an output label; learning this mapping requires a collected set of training data with known labels. We will use data from MIMIC-III critical care database, decision tree and logistic regression in our project. We will also use moving average to show the real health condition of a patient. We also have a plan to use neural network algorithm in this project. Here are the details of all methodologies and methods of our thesis project.

2.1 ICU Patient monitoring system

Patients admitted to acute hospitals today are sicker than in the past, as they have more complex health problems and are far more likely to become seriously ill during their admission. In addition, patients who were once too sick to be operated on are now undergoing complex surgical procedures. After that they pass a very critical period. They need very close observation. Some parameters are needed to be checked in every minute. There are dedicated machines in ICU unit for this observation. Here is eight major parameters for every ICU patient to be monitored always.

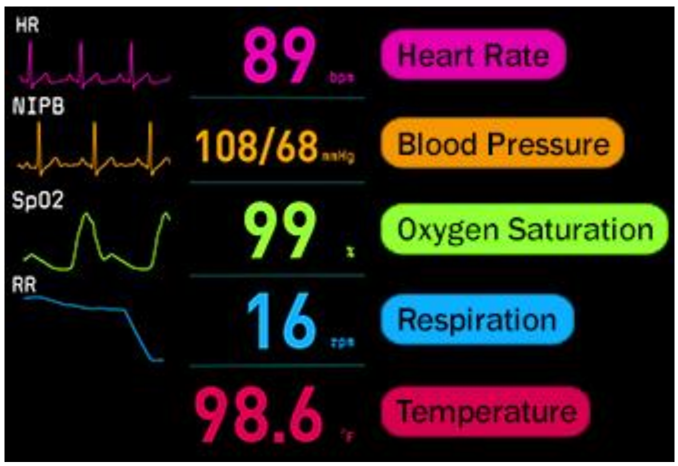

Figure 1. ICU parameters (Source: Courtesy to Phillips).

2.1.1 Blood Temperature

The body's temperature represents the balance between heat produced and heat lost otherwise known as thermoregulation. In the clinical environment, body temperature may be affected by factors such as underlying pathophysiology (e.g. sepsis), skin exposure (e.g. in the operating theatre) or age [7]. Other factors may not affect the body's core temperature but can contribute to inaccurate measurements, such as the consumption of hot or cold fluids prior to oral temperature measurement [9]. Clinically, there are three types of body temperature: the patient's core body temperature; how the patient says they feel; and the surface body temperature or how the patient feels to touch. Lastly, we had to consider the temperature of their blood.

Importantly, these three are not always the same and may differ according to the underlying disease process. We need the blood temperature in ICU. There are some patients who faces severe trauma if their blood temperature is low [41]. This temperature is always monitored in ICU.

2.1.2 Pulse

Pulse is defined as the palpable rhythmic expansion of an artery produced by the increased volume of blood pushed into the vessel by the contraction and relaxation of the heart [29]. The pulse is affected by many factors including age, existing medical condition, medications (e.g. betablockers) and fluid status (e.g. hyper/ hypovolaemia). Nurses should be aware that the pulse is not always a true reflection of cardiac contractility or output; in the case of aortic stenosis, for example, the pulse may be weak in spite of forceful cardiac contractions [40]. Pulse should also not be considered the same as heart rate, which is actually a measurable pulse characteristic. When the pulse is palpated, characteristics other than rate also need to be assessed. These include the strength or amplitude of the pulse, peripheral equality of pulses and the pulse's regularity. In ICU pulse rate is always monitored by a sensor clipping on the finger and it is shown on the monitor [25].

2.1.3 Blood pressure

Blood pressure (BP) refers to the pressure exerted by blood against the arterial wall. It is influenced by cardiac output, peripheral vascular resistance, blood volume and viscosity and vessel wall elasticity [12]. BP is an important vital sign to measure as it provides a reflection of blood flow when the heart is contracting (systole) and relaxing (diastole). It is also one of many indicators of cellular oxygen delivery. Changes or trends in BP may reflect underlying pathophysiology or the body's attempts to maintain homeostasis [8]. A drop in BP, for example, has been found to be a common sign in patients prior to cardiac arrest. A change in BP alone, though, does not indicate that the patient will have a cardiac arrest, but should trigger the nurse to perform a more detailed assessment. The importance of measuring BP accurately cannot be over-emphasized; and yet, it is one of the most inaccurately measured vital signs. If a BP reading consistently underestimates the diastolic pressure by 5mm Hg, it could result in two thirds of hypertensive patients being denied preventative treatment[25][31]. SO, blood pressure is very important for monitoring. In ICU blood pressure is always monitored and shown in a monitor.

2.1.4 Respiratory rate

Respiratory rate is an important baseline observation and its accurate measurement is a fundamental part of patient assessment. Respiratory rate measurement serves a number of purposes, such as being an early marker of acidosis [47]. It is also one of the most sensitive indicators of critical illness. An increase from the patient's normal rate of even three to five breaths per minute is an early and important sign of respiratory distress and potential hypoxemia. Despite this, research has found that the respiratory rate is often not recorded in clinical settings or is simply guessed [16]. This is disturbing given that an abnormal respiratory rate is the best predictor of an impending adverse event such as cardiac arrest. The reason for this haphazard assessment is unclear. Perhaps it is because nurses assume that oxygen saturation provides a greater reflection of the patient's respiratory function or because there is no automated machine for measuring respiratory rate [28]. In the acutely ill patient, respiratory rate should be counted for a full minute, rather than 30 seconds and then doubled. In measuring the respiratory rate, the pattern should also be assessed and classified as eupnea, tachypnea, bradypnoea or hypopnea [32]. Labeling it as such encourages the nurse to do more than simply count a number, and to consider why the respiratory rate may be fast or slow. Nurses should also assess respiratory effort (depth of inspiration and use of accessory muscles) and equality of thoracic expansion.

2.1.5 Urine output

Urine output is an indirect reflection of renal function and fluid status and, therefore, should be monitored closely in acutely ill patients [25]. Urine output is not an absolute indicator of renal failure as such; but rather, may be the first clinical indicator of a fluid and electrolyte imbalance, which if left untreated may lead to renal failure. In a study of over 2000 consecutive hospital admissions, 5% of patients developed acute renal failure, with the main causes being hypo perfusion (e.g. hypotension, cardiac dysfunction) and major surgery. In hospital-acquired acute renal failure, hypo perfusion is the most common cause. The patients most likely to develop acute renal failure are those with poor renal perfusion, pre-existing renal dysfunction, diabetes mellitus, sepsis, vascular disease and liver disease [19].

2.1.6 Oxygen saturation (SpO2)

A pulse oximeter is an extremely valuable clinical tool and easy to use [43]. Research suggests that pulse oximetry is useful for detecting a change in condition that may otherwise have been missed, resulting in changed patient management and a reduction in the number of investigations undertaken. To avoid error with its use, the nurse must understand the factors that affect its accuracy, but studies have shown this knowledge is often lacking [41]. Specifically, the nurse must understand respiratory physiology, how to assess peripheral circulation and the oxyhemoglobin dissociation curve. To work effectively, a pulse oximeter requires an adequate peripheral blood flow. This flow may be impaired by factors such as patient movement (e.g. tremor, shivering), hypovolemia, hypothermia, arrhythmias, vasoconstriction or heart failure. The SpO2 reading may also be misleading if the patient is anemic, because an oximeter does not measure the patient’s hemoglobin level, and an anemic patient may have a normal SpO2 despite having a lowered potential to carry oxygen [11]. For this reason, we did not rely on SpO2 as the sole indicator of the patient’s tissue oxygenation.

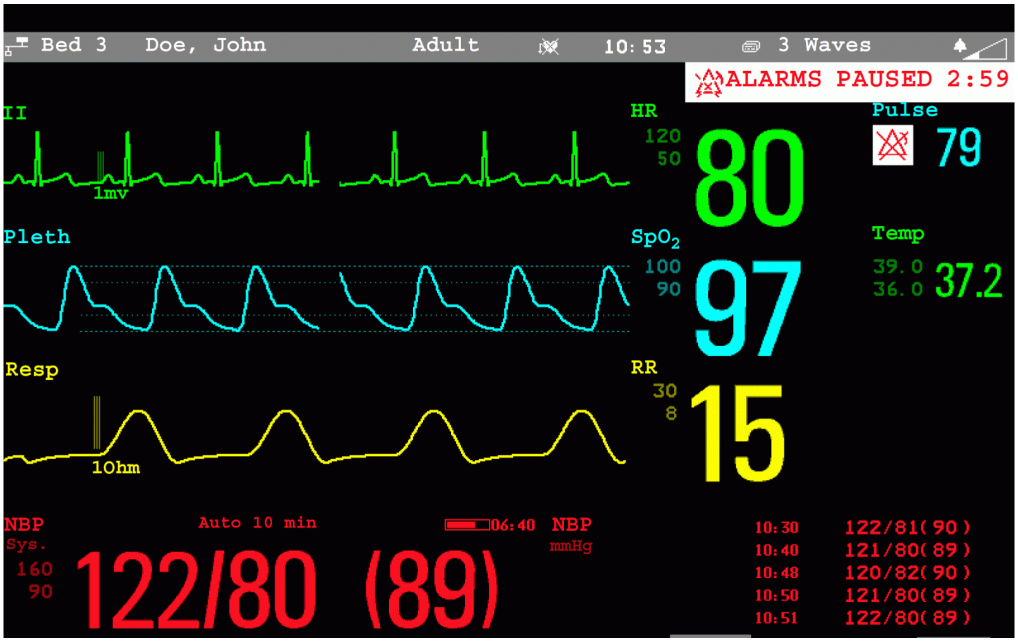

Figure 2. ICU monitor (Source: Courtesy to Phillips).

2.2 Data collection from MIMIC-III database

Data collection is very important part for our project. Before describing the process of data collection we have to know what kind of real time data an ICU patient produces and what are the systems for monitoring ICU patient. Then we will know what kind of data is stored in database like MIMIC-III. After that we will describe why we choose our needed parameter between huge ranges of parameter collections and how we made our own dataset from MIMIC-III database.

MIMIC-III (‘Medical Information Mart for Intensive Care’) is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital. Data includes vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, and diagnostic codes, imaging reports, hospital length of stay, survival data, and more. The database supports applications including academic and industrial research, quality improvement initiatives, and higher education coursework. This database includes more than 20,000 patients admitted to various intensive care units (ICUs) (e.g., medical, surgical, coronary care, and neonatal) of Beth Israel Deaconess Medical Center (BIDMC, Boston, MA, USA) between 2001 and 2008[49]. The Institutional Review Boards (IRB) of the Massachusetts Institute of Technology (MIT, Cambridge, MA, USA) approved the establishment of this database. All patients in this database are de-identified, and key demographics (e.g., admitted time, birthday) are shifted to protect their privacy [49].

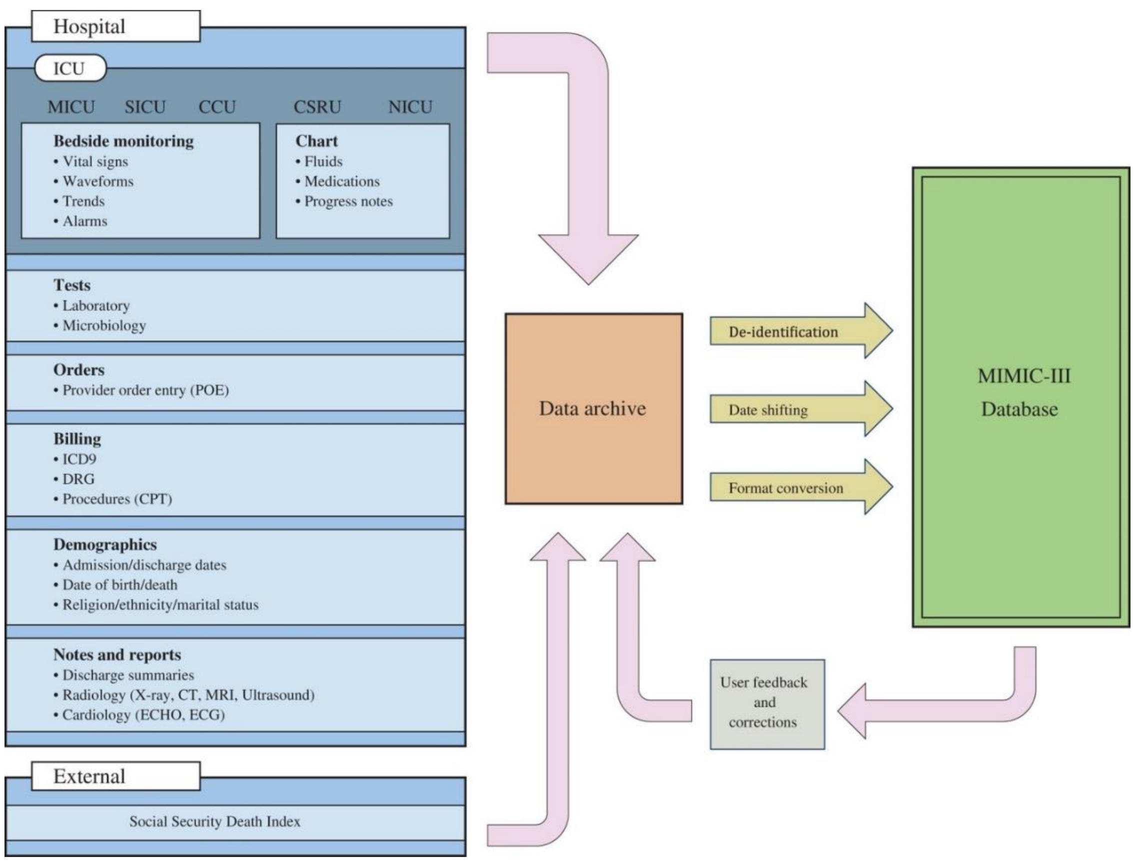

Figure 3. MIMIC -III database structure (Source: Physionet.com).

MIMIC-III is a relational database consisting of 26 tables. It is provided as a collection of comma separated value (CSV) files, along with scripts to help with importing the data into database systems including PostreSQL, MySQL, and

MonetDB [49]. As the database contains detailed information regarding the clinical care of patients, it must be treated with appropriate care and respect. We are using this in our project. It has a huge dataset and huge collection of patient data. As a result we can use it in our project for more accurate results.

MIMIC has been used as a basis for coursework in numerous educational institutions, for example in medical analytics courses at Stanford University (course BIOMEDIN215), Massachusetts Institute of Technology (courses HST953 and HST950J/6.872), Georgia Institute of Technology (course CSE8803), University of Texas at Austin (course EE381V), and Columbia University (course G4002), amongst others[49]. MIMIC has also provided the data that underpins a broad range of research studies, which have explored topics such as machine learning approaches for prediction of patient outcomes, clinical implications of blood pressure monitoring techniques, and semantic analysis of unstructured patient notes. There are two main types of data in MIMIC [49]. One is numerical data and another one is wave form of data. Figure 4 shows sample wave data of MIMIC for a single patient stay in a medical intensive care unit. The patient, who was undergoing a course of chemotherapy at the time of admission, presented with febrile neutropenia, anemia, and thrombocytopenia [3].

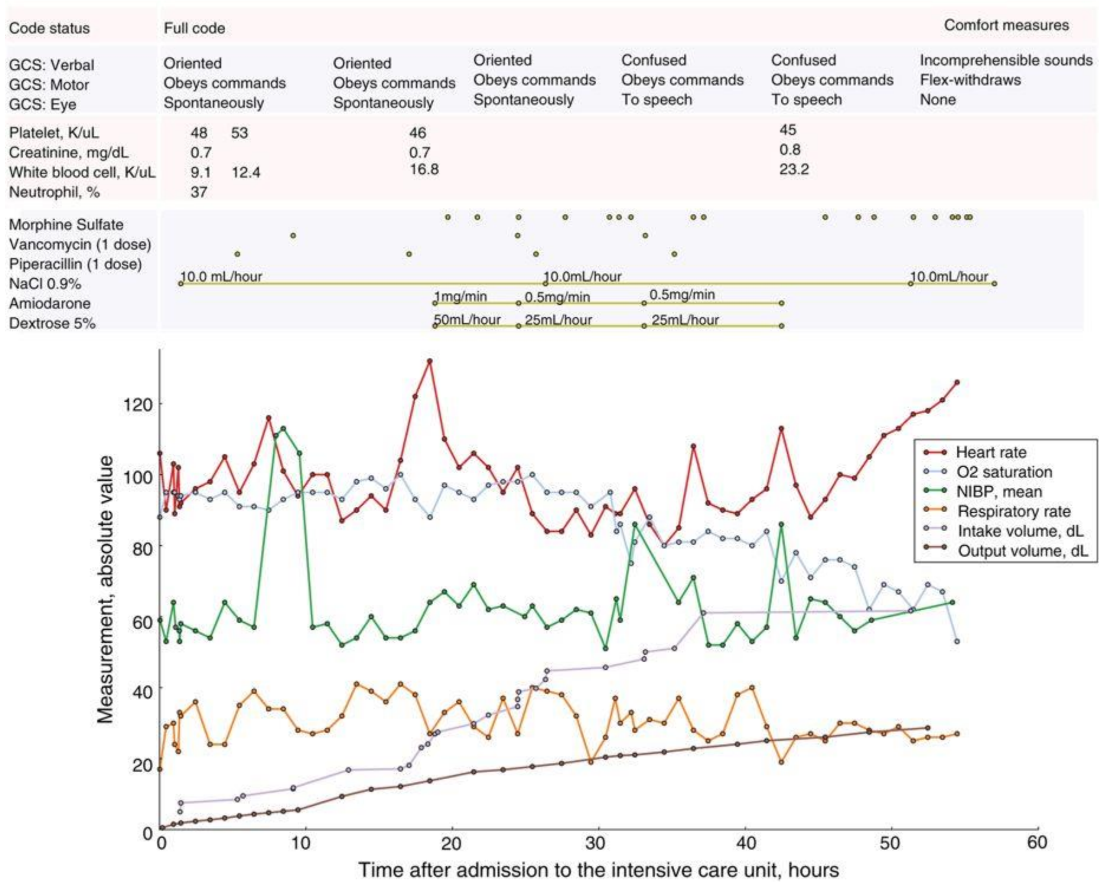

Figure 4. Sample data for single patient stay in a medical intensive care unit (Source: MIMIC-III, a freely accessible critical care database.

2.3 Decision tree

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too[46].The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data (training data). The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label. Decision trees used in data mining are mainly of two types [11]:

1. Classification tree in which analysis is when the predicted outcome is the class to which the data belongs. For example, outcome of loan application are safe or risky.

2. Regression tree in which analysis is when the predicted outcome can be considered a real number. For example, the population of a state.

Both the classification and regression trees have similarities as well as differences, such as procedure used to determine where to split. There are various decision trees algorithms namely ID3 (Iterative Dichotomizer 3), C4.5, CART (Classification and Regression Tree), CHAID (Chi - Squared Automatic Interaction Detector), MARS. Out of these, we will be discussing the more popular ones which are ID3, C4.5, and CART. We are using ID3 mainly in our thesis. We have also tested C4.5 for accuracy checking.

2.2.1 Advantages of ID3

I. The training data is used to create understandable prediction rules.

II. It builds the fastest as well as a short tree.

III. ID3 searches the whole dataset to create the whole tree.

IV. It finds the leaf nodes thus enabling the test data to be pruned and reducing the number of tests.

V. The calculation time of ID3 is the linear function of the product of the characteristic number and node number.

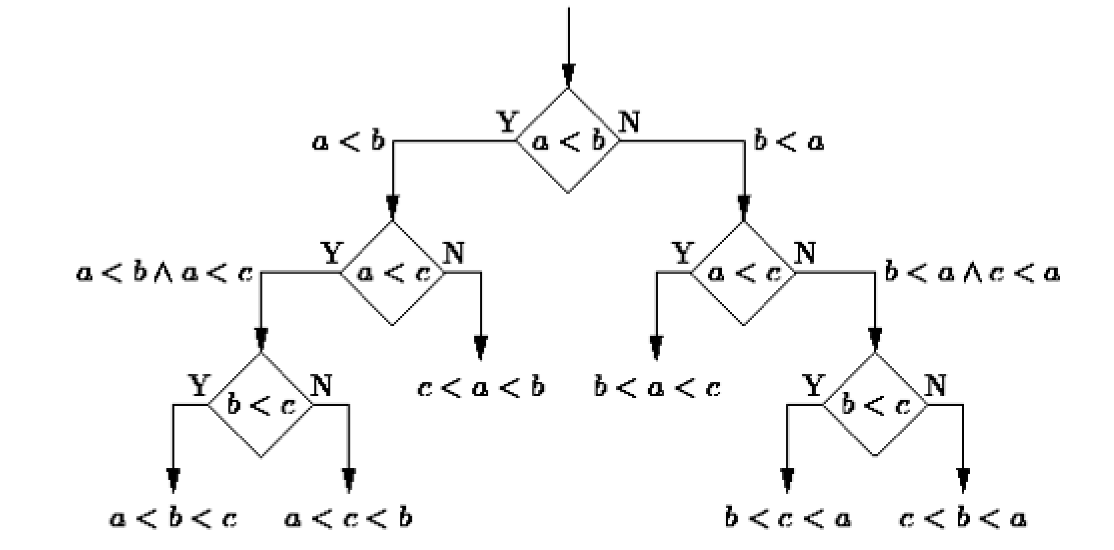

Figure 5. Basic construction of a decision tree (Source: Machine learning directory.com).

Decision Tree Algorithm Pseudocode:

I. Place the best attribute of the dataset at the root of the tree.

II. Split the training set into subsets.
III. Subsets should be made in such a way that each subset contains data with the same value for an attribute.

Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

In decision trees, for predicting a class label for a record we start from the root of the tree [29]. We compare the values of the root attribute with record’s attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node. We continue comparing our record’s attribute values with other internal nodes of the tree until we reach a leaf node with predicted class value. As we know how the modeled decision tree can be used to predict the target class or the value. Now let’s understanding how we can create the decision tree model.

2.3.2 Decision tree working process:

1. Attribute selection
2. Information gain

Attributes Selection:

If dataset consists of "n" attributes then deciding which attribute to place at the root or at different levels of the tree as internal nodes becomes a complicated step. By just randomly selecting any node to be the root can't solve the issue. If we follow a random approach, it may give us bad results with low accuracy [21].

For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some criterion like information gain, gini index, etc. [21]. These criterions will calculate values for every attribute. The values are sorted, and attributes are placed in the tree by following the order i.e. the attribute with a high value (in case of information gain) is placed at the root. While using information Gain as a criterion, we assume attributes to be categorical, and for gini index, attributes are assumed to be continuous.

Example:

If S is a collection of 14 examples with 9 YES and 5 NO examples then

Entropy(S) = - (9/14) Log2 (9/14) - (5/14) Log2 (5/14) = 0.940 (1)

Notice entropy is 0 if all members of S belong to the same class (the data is perfectly classified). The range of entropy is 0 ("perfectly classified") to 1 ("totally random").

Gain(S, A) is information gain of example set S on attribute A is defined as

Gain(S, A) = Entropy(S) - S ((|Sv| / |S|) * Entropy (Sv)) (2)

Where-

S is each value v of all possible values of attribute A

Sv = subset of S for which attribute A has value v

|Sv| = number of elements in Sv

|S| = number of elements in S

Information Gain:

By using information gain as a criterion, we try to estimate the information contained by each attribute. We are going to use some points deducted from information theory. To measure the randomness or uncertainty of a random variable X is defined by Entropy.

For a binary classification problem with only two classes, positive and negative class.

I. If all examples are positive or all are negative then entropy will be zero i.e. low.

II. If half of the records are of positive class and half are of negative class then entropy is one i.e. high.

By calculating entropy measure of each attribute we can calculate their information gain. Information Gain calculates the expected reduction in entropy due to sorting on the attribute. Information gain can be calculated. To get a clear understanding of calculating information gain & entropy, we will try to implement it on our sample data.

2.3Logistic Regression

Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function [8]. Let's understand the above logistic regression model definition word by word. What logistic regression model will do is, it uses a black box function to understand the relation between the categorical dependent variable and the independent variables. This black box function is popularly known as the Softmaxfunciton. The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables [14]. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k \qquad (3)$$

Here p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic} \quad (4)$$

Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values. Sometimes logistic regressions are difficult to interpret; the Intellectual Statistics tool easily allows you to conduct the analysis, and then in plain English interprets the output.

2.4.1 Type of questions that a binary logistic regression can examine.

How does the probability of getting lung cancer (yes vs. no) change for every additional pound a person is overweight and for every pack of cigarettes smoked per day?

Do body weight, calorie intake, fat intake, and age have an influence on the probability of having a heart attack (yes vs. no)?

We are using logistic regression in our thesis to control the ICU machines. It will use two or three parameter to decide what to do. When and how the machines will react, this decision is taken by logistic regression.

2.4.2 Over fitting.

When selecting the model for the logistic regression analysis, another important consideration is the model fit. Adding independent variables to a logistic regression model will always increase the amount of variance explained in the long odds (typically expressed as R^2). However, adding more and more variables to the model can result in over fitting, which reduces the generalizability of the model beyond the data on which the model is fit.

Proposed model for Improving ICU decision support

When the model starts, it takes training dataset as input. Then it checks the datasets. After checking it cleans the dataset and handle if any data is missing.

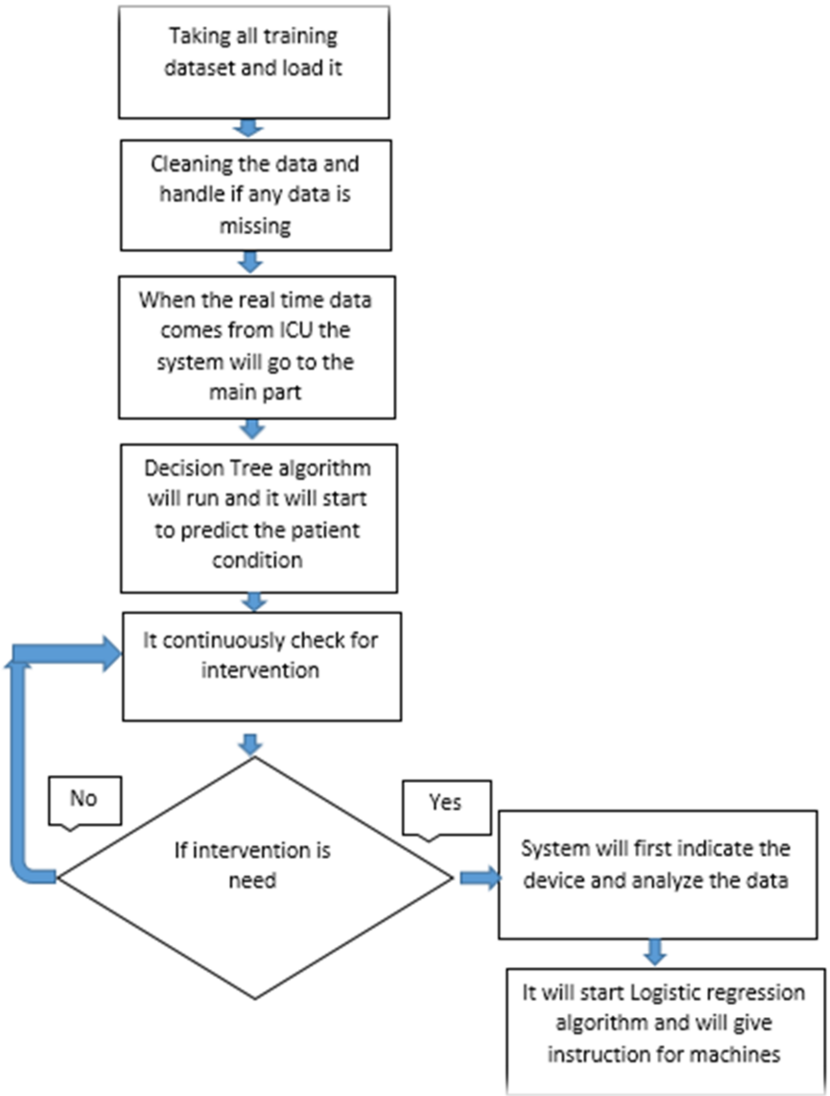

Figure 6. Work flow of proposed model.

Dataset checking is important because if there is anything wrong with the dataset, the answer will not be accurate. After checking dataset the model will take real time data input from ICU. Then the model steps into algorithm part. After getting real time ICU data the decision tree will start to predict the future condition of the patient. It shows a curve to indicate risk. The model always check for intervention part. When there is any input of unusual data(less than threshold level), it will use logistic regression algorithm to calculate the decision. Logistic regression gives result in binary (yes or no). If the answer is yes the machine will turn on otherwise it will remain off.

3.1 Data preparation for models

MIMIC-III is a huge database. It has over 144 thousand patient's data but we do not need that much data for our thesis. We will collect 500 data sample for our thesis and make our own dataset. MIMIC-III has 26 tables of its own but in our thesis we will take data from only three tables. We are directly taking those tables. We are collecting rows from those three tables and making our own dataset. We are taking some rows for our system and in real ICU system there will be other data tables. When we will implement our system in real ICU, we will establish relationship between those tables with our ones. The data we take from MIMIC-III is basically for our training data set. We will train our system with our dataset.

In main ICU dataset there is column of patient id, patient sex, patient age and patient ward no and unit number is included as table1.

Table 1. Patients Identity data sheet.

SL#	ID	Sex	Age	Ward No	Unit no
1	00012	Male	54	2	8
2	00143	Female	67	4	2
3	98912	Female	39	8	9
4	27127	Male	17	1	3

The datasets that we made for us as our train dataset has 7 columns (Table 2) a, 4(Table 3) and columns 3 columns (table 4). These dataset we used for our machine learning algorithm. Table 2 is used for controlling the ventilation machine and syringe pump. Table 3 is used for blood warmer.

Table 2. Train dataset 1 sample.

ID	B.Sgr	BP	Pulse rate	Spo_2	Herat beat	Ventilation

1.	7.93	60	30	75	86	On
2.	11.5	75	70	96	72	OFF
3.	15.7	65	40	80	88	On
4.	8.0	85	65	97	68	Off
5.	17.67	80	71	96	65	Off
6.	12.40	85	75	98	76	Off

Table 3. Train dataset for controlling blood warmer machine.

Sl#	Mean of temp 4hr(c)	Temp present(c)	Blood warm
1.	34.65	32	on
2.	35.5	36	off
3.	35.87	31.5	on
4.	36.78	37	off
5.	37.43	35.5	off

Table 4. Train Data set to operate syringe pump.

Sex	Kreatine	Sugar	Pump
Male	.72	7.8	off
Female	.8	8.9	off
Female	1.25	13.7	On
Male	.7	6.7	off

3.2 Working logic and implementing algorithms

3.2.1 Risk prediction for heart patient

When a patient is admitted in ICU he/she has an identity data at that moment. All patients' identity data is kept in main dataset (table 1) but there is also different datasets which contains different values for different type of people. When this proposed model predicts future for an ICU patient it needs four basic parameters. One of them is unit number. This number indicates the injury type of patient. Here, we have predicted risk and future of heart patient. After checking its ward this proposed model check there is any surgery or not. Depending on the surgery decision the risk varies. Suppose when a patient has heart operation his heart beat range is between 90-110bmp [13]. One the other hand when the patient does not have any surgery in heart the heart beat rate is 82-89bmp [13]. This model will show a time series for both of the patient to understand the risk clearly. There is one curve for the prediction. That comes from the train dataset and another one is from the real time patients data. In real time data record of heart beat rate is kept hourly. First model will task dataset1 as train dataset (Figure 7).

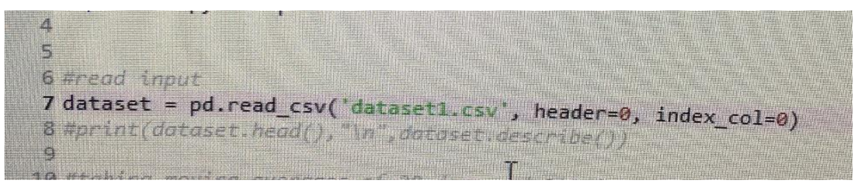

Figure 7. Giving the training dataset1 in model.

In figure 8. We can observe the time is horizontally and the heart beat rate is in vertically. Here the red line is coming from the training dataset. This line shows us the risk level. Here we can observe in 9th hour there is less heart rate. This is because of medicine. 10th – 12th hour is very critical for heart patient [18]. So then a medicine is given for keeping the heart rate low [18]. As it is low so if it increases there is less chance of death. The blue one is for showing the patient's condition.

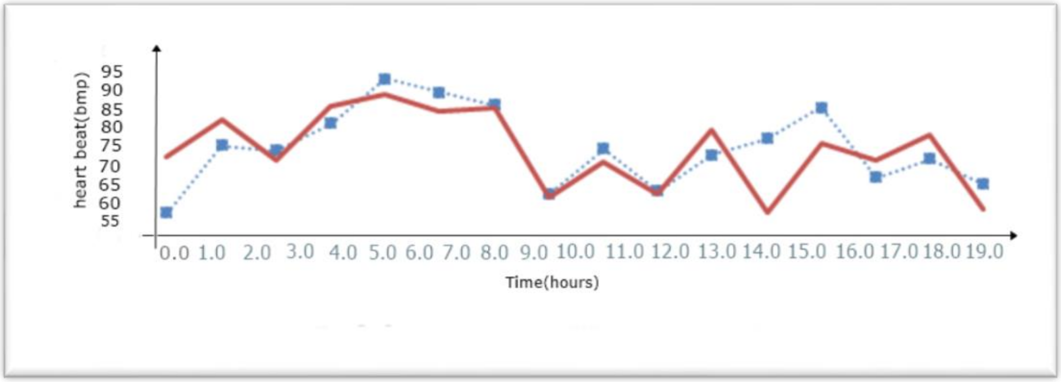

Figure 8. Showing the relation between the training dataset of heart rate and the real time patient's heart rate to observe the change directly.

After this the model will show the patient is going at risk or improving. For calculating the risk the proposed model will count 4 things. Firstly, it will check the sugar level of the patient. Sugar level is important because if the patient has diabetics he/she can need some more medication to recover the operation. After checking the sugar it will check pulse. Then it will check the spo_2 and respiration rate. There is a certain threshold level for every parameter from the train dataset. The model will check one by one and every decision is important here. One decision is related with another parameter's decision. As a result this whole checking is done by the decision tree.

3.2.2 Implementation of ID_3 Algorithms

For this implementation Numpy arrays and pandas data frames will help us in manipulating data. Sklearn is a machine learning library. The tree module will be used to build a Decision Tree Classifier for predicting the risk. Accuracy score module will be used to calculate accuracy metrics from the predicted class variables. After putting the dataset at the same folder, we will use Pandas read_csv() method to import data into pandas data frame. Since our real time data is separated by commas "," and there is no header in our data, so we will put header parameter's value "None" and sep parameter's value as ",". The data import part is done for the test data set now we also have to load the training data set. Then Time to run the decision tree algorithm. We will make two classifier (Figure 9) then we will predict the future condition of the patient by those. The prediction part is dependent on decision tree classifier part. This is the classifier function for Decision Tree. It is the main function for our model for implementing the algorithm. Some important parameters are:

Criterion: It defines here the function to measure the quality of a split. Sklearn supports “gini” criteria for Gini Index & “entropy” for Information Gain.

Splitter: It defines the strategy to choose the split at each node. Supports “best” value to choose the best split & “random” to choose the best random split. By default, it takes “best” value. For example, in this model when it chooses the heart beat rate for non-surgical patient, it will split at value 60.

max_features: It defines the no. of features to consider when looking for the best split. We can input integer, float, string & none value. Here we have used integer, float and string for our proposed model.

```
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=3,
            max_features=None, max_leaf_nodes=None, min_samples_leaf=5,
            min_samples_split=2, min_weight_fraction_leaf=0.0,
            presort=False, random_state=100, splitter='best')
```

Figure 9. Classifier method of our proposed model.

If an integer is inputted then it considers that value as max features at each split. If float value is taken then it shows the percentage of features at each split. If “auto” or “sqrt” is taken then max_features=sqrt(n_features). If “log2” is taken then max_features = log2 (n_features). If none, then max_features = n_features. By default, it takes “None” value.

max_depth: The max_depth parameter denotes maximum depth of the tree. It can take any integer value or none. If none, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples. By default, it takes “None” value.

min_samples_split: This tells above the minimum no. of samples reqd. to split an internal node. If an integer value is taken then consider min_samples_split as the minimum no. If float, then it shows percentage. By default, it takes “2” value.

min_samples_leaf: The minimum number of samples required to be at a leaf node. If an integer value is taken then consider min_samples_leaf as the minimum no. If float, then it shows percentage. By default, it takes “1” value.

max_leaf_nodes: It defines the maximum number of possible leaf nodes. If none then it takes an unlimited number of leaf nodes. By default, it takes “None” value.

min_impurity_split: It defines the threshold for early stopping tree growth. A node will split if its impurity is above the threshold otherwise it is a leaf.

After getting two classifiers for gini index and information gain we can use them to predict the future of the patient. Here below is clear idea about Entropy and information gain.

Entropy

In our model decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one.

To build our desired decision tree, we had to calculate entropy with two attributes using frequency tables as follows:

a. Firstly we have calculated entropy of one attributes using this formula.

b. Then we have find we used to find Entropy using the frequency table of two attribute by using this formula.

Suppose, our system is calculating the health condition of the patient but Patient’s health is improving or not. It will take two attributes

E(Health_condition,present_rate)=
P(Pulse)*E(yes,no)+P(Spo$_2$)*E(yes,no)+P(RR)*E(yes,no)+P(Heart rate)*E(Yes,no)

Here, Yes and no values come from table depending on Real time values.

Information Gain

Our information gain is based on the decrease in entropy after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain.

First we have to calculate the entropy of the target. The dataset is then split on the different attributes. The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split. The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy.

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X) \qquad (5)$$

Then, Choosing attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch. A branch with entropy of 0 is a leaf node. A branch with entropy more than 0 needs further splitting. The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified. After that it can give final decision about the patient's health.

3.2.3 Methods of Controlling the Machines of ICU

In our proposed model we had a plan to control the important machines of ICU. Machines like ventilation machine, blood warmer machines and Syringe pump. For controlling these machines first we had to set the level when the machine will start by itself and when it will stop operating. For setting this level we had keep in mind three things. First one is Spo_2 rate. Second one is RR rate and then the pulse. Ventilation is needed when the lungs cannot take air. So the Spo_2 goes down and as result the RR rate goes down and pulse rate also goes down. Pulse rate can goes down for other reasons too and RR rate can also goes down for other reasons. When one and only the spo_2 goes down and those two rates also goes down, the ventilation will start. Though they are internally dependent, they are independent variable in this proposed model. They have the same importance. As they are Independent we decided to take this decision by logistic regression. After taking the decision we will save it in a variable and then we will compare it with one previously plotted variable. If logistic regression gives positive value our model will turn that on and otherwise it will remain off. This same formula will also work for blood warmer and syringe pump machines. There the dependent and independent variable will be different.

Logistic Regression is a Machine Learning algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression predicts will predict in our model the P(Y=1) as a function of X. Here is implementation of this method in our model.

3.2.4 Implementation of logistic regression:

In our model we have to do this method twice. One is for our training dataset and another for our real time patient's dataset. Then we compare those two for final decision.

Input variable: x- SPO_2, RR rate, pulse rate

Predict variable (desired target): y- Does the patient need ventilation? (Binary: "1", means "Yes", "0" means "No")(Figure 10)

The observed probability of Y=1 for each level of X, calculated as the ratio of the number of instances of Y=1 to the total number of instances of Y for that level. The odds for each level of X, calculated as the ratio of the number of Y=1 entries to the number of Y=0 entries for each level, or alternatively as observed probability/(1 - observed probability). The natural logarithm of the odds for each level of X, designated as "log odds". In this proposed method mechanics of the process begin with the log odds, which will be equal to 0.0 when the probability in question is equal to .50, smaller than 0.0 when the probability is less than .50, and greater than 0.0 when the probability is greater than .50.

```
#Data setup

Species = ['Spo2', ' RR_rate', 'Pulse_rate']

#Number of examples
m = Ventilation_data.shape[0]
#Features
n = 3
#Number of classes
k = 3

X = np.ones((m,n + 1))
y = np.array((m,1))
X[:,1] = Ventilation_data['Spo2'].values
X[:,2] = Ventilation_data['RR_rate'].values
X[:,3] = Ventilation_data['Pulse_rate'].values

#Labels
y = Ventilation_data['Ventilation'].values

#Mean normalization
for j in range(n):
    X[:, j] = (X[:, j] - X[:,j].mean())

X_train, y_train,train_test_split(X, y, test_size = 0.2, random_state = 11)
```

Figure 10. Data setup part of logistic regression code. In this part the data coming from training set is fitted.

The form of logistic regression supported by the present page involves a simple weighted linear regression of the observed log odds on the independent variable X(spo2,RR,pulse).

When the model predicts for the blood warmer machine and syringe pump then it has to concentrate only two independent variables. The dependent variable is same. The on/off decision is the dependent variable.

```python
#Logistic Regression

def sigmoid(z):
    return 1.0 / (1 + np.exp(-z))

#Regularized cost function
def regCostFunction(theta, X, y, _lambda = 0.1):
    m = len(y)
    h = sigmoid(X.dot(theta))
    reg = (_lambda/(2 * m)) * np.sum(theta**2)

    return (1 / m) * (-y.T.dot(np.log(h)) - (1 - y).T.dot(np.log(1 - h))) + reg

#Regularized gradient function
def regGradient(theta, X, y, _lambda = 0.1):
    m, n = X.shape
    theta = theta.reshape((n, 1))
    y = y.reshape((m, 1))
    h = sigmoid(X.dot(theta))
    reg = _lambda * theta /m

    return ((1 / m) * X.T.dot(h - y)) + reg

#Optimal theta
def logisticRegression(X, y, theta):
    result = op.minimize(fun = regCostFunction, x0 = theta, args = (X, y),
                         method = 'TNC', jac = regGradient)

    return result.x
```

Figure 11. Logistic regression part of the controlling model.

4. Result and discussions

In proposed model we have used python as programming language because python is efficient for machine learning algorithms as we used machine learning algorithms. We tried to measure the health condition of ICU patient in proposed model. Proposed model can predict the future condition of ICU patient based on its huge dataset. It can also control some lifesaving machines. It will give signal ventilation machine to operate id the spo_2 and RR rate is low. Then it also can give signal to blood warmer and the syringe pump. ICU has huge machineries and we had a limited time and resource to do this task. So far we ended up with three machines.

We made total 4 datasets by collecting data from MIMIC-III database. Our datasets are leveled data and the test dataset is separated by the comma. This model firstly gives a clear view of heart rate in comparison with a standard result prepared by the train dataset. Then it calculates the risk. The risk is calculated on the basis of 4 given parameters (Spo_2, RR_rate, Pulse_rate,Bp). The result shows us the future condition of the patient. We also plotted the data after prediction (Figure 13).

```
array(['U', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L',
       'L', 'U', 'L', 'L', 'L', 'L', 'L', 'L', 'U', 'L', 'L', 'U', 'L',
       'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'U', 'L',
       'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'U', 'L', 'U', 'U',
       'L', 'L', 'L', 'L', 'U', 'L', 'L', 'L', 'U', 'U', 'L', 'L', 'L',
       'L', 'L', 'L', 'L', 'L', 'L', 'L', 'U', 'L', 'U', 'L', 'U', 'U',
       'L', 'L', 'L', 'L', 'L', 'L', 'L', 'U', 'U', 'L', 'L', 'U', 'U',
       'L', 'L', 'L', 'L', 'U', 'L', 'L', 'L', 'L', 'U', 'L', 'U', 'U',
       'L', 'L', 'U', 'L', 'U', 'L', 'L', 'L', 'U', 'U', 'L', 'U', 'U',
       'L', 'L', 'U', 'L', 'U', 'L', 'L', 'L', 'L', 'L', 'L', 'L', 'U',
       'L', 'L', 'U', 'L', 'U', 'L', 'U', 'L', 'L', 'U', 'L', 'U', 'L',
       'U', 'L', 'L', 'L', 'L', 'L', 'U', 'L', 'U', 'L', 'L', 'U', 'U',
       'L', 'U', 'U', 'U', 'U', 'U', 'L', 'L', 'U', 'L', 'L', 'U', 'L',
       'L', 'U', 'U', 'L', 'U', 'U', 'L', 'L', 'L', 'L', 'L', 'L', 'U',
       'U', 'L', 'L', 'U', 'U', 'L'], dtype=objec
```

Figure 12. Output of the risk prediction. Here U means upper than the risk level and L means lower than the risk level.

The result of figur12 shows us only the patient when can be in risk. We constructed a curve to indicate the risk line from the training dataset and the result of figure 11. In that graph score 1 indicates the full risky situation with red line and with green line it indicates the current patients situation. If the green line is much closer to red line that means the patient moving into risk zone. The red line indicates death. When a patient is dead then real time data of heart beat will be 0. Then at first stage of calculating decision tree it will eliminate. As a result the green will cross the level of will touch the level. If a patient's heart beat is in good condition but spo_2 is very low then the green curve will go near to the red one but it will touch as living human's Spo2 rate never goes near 0. We collected the result from decision tree then saved it in a notepad. Using that result this Figure 12 is plotted.

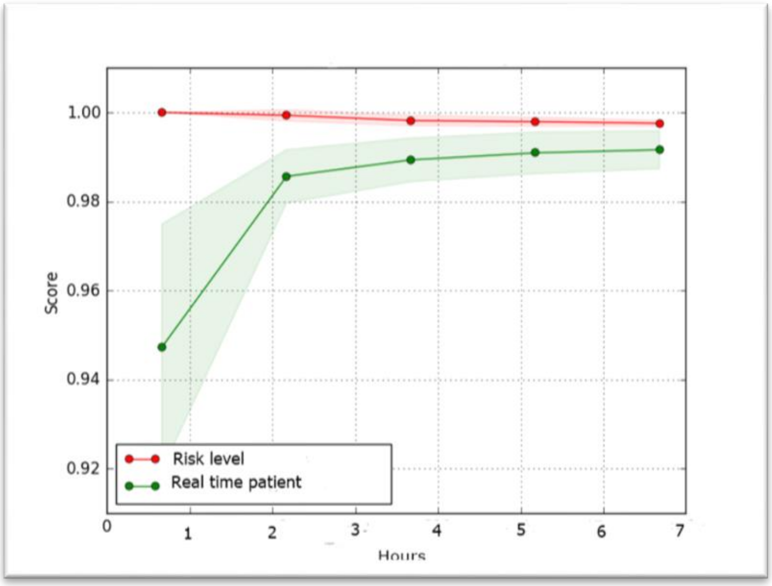

Figure 13. Risk prediction curve (After collecting data from decision tree's answer, we plotted them in graph directly. We used only 6 hours interval where 0 in not counted. So plotting starts from .7 and stops at 6.7.).

When our proposed model starts analyze data for controlling machines it uses logistic regression. In this case results came depending on independent variables. Analyzing the result we understood in case of blood warmer machine the dependent variables are temperature and the mean of the previous 4 hours temperature. There the model compares present temperature with previous mean temperature. After that it gives binary answer of 1 or 0. Then it checks the training dataset. It searches for the similar condition. Furthermore, based on that matching it gives final output of true or false. Where true means blood warmer machine will turn on itself and false means it will not turn on. We can see a large result list where yes means on and no means off (Figure 14).

This proposed model also gives solution to control syringe pump using logistic regression. In this case the independent variables are blood sugar and kreatine and

```
[False False False False  True False False False  True False False  True
 False False False  True False  True  True False False False False False
 False False False False False False False False  True False False False
 False False False False False False  True  True  True False False False
  True  True  True False False False  True False False  True  True  True
  True]
```

Figure 14. Logistic regression's final binary answer.

dependent variable is the decision. Blood sugar and the kreatine rate is independent just because of their own value. They can change the health condition without depending on anything. So, the answer is yes when the blood sugar rate and the kreatine rate is high. For example, when the blood sugar rate is high it will automatically inject insulin into patient's body. There is no need of nurses to check. We tried this several times with changing dataset but the current dataset is giving the most accuracy (Figure 15).

```
#Predictions
P = sigmoid(X_test.dot(all_theta.T))
p = [Species[np.argmax(P[i, :])] for i in range(X_test.shape[0])]

print("Test Accuracy ", accuracy_score(y_test, p) * 100 , '%')
```

```
Test Accuracy  96.66666666666667 %
```

Figure 15. Accuracy checking of Logistic regression.

The proposed model will show also a moving cross over average of patient health condition. On x-axis of that will be the hours and y axis there will be the health fitness rate. This health fitness rate is the data that we got previously in decision tree. The proposed model will just plot it.

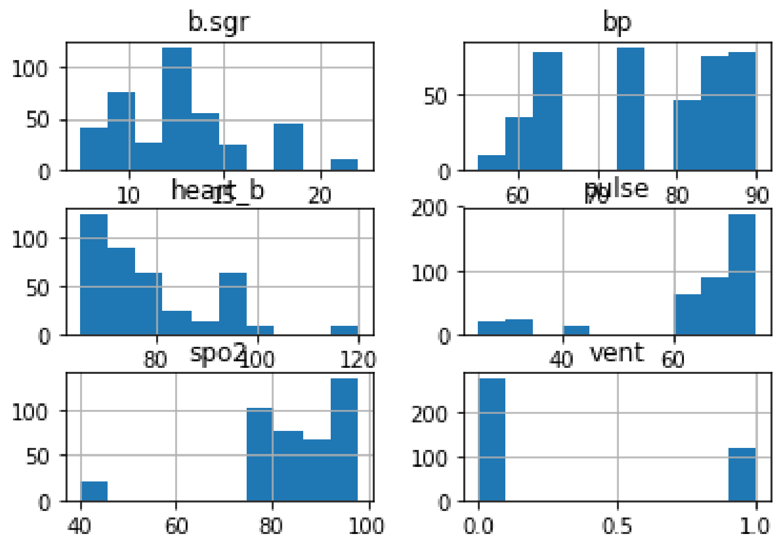

Figure 16. Comparison of other data with ventilation

In Figure 16 we can see the comparison of other data with ventilation. We get this result from logistic regression.

5. Conclusion

The proposed model not only shows the future health condition but also it will automate some ICU machines. It also observes the risk on ICU patient's health. The main problem with this kind of research work is the lacking of proper data even though we managed dataset from MIMIC-III database. There are four main training dataset in the proposed model. Most importantly, decision tree gives the prediction of health condition and logistic regression decides when intervention is needed. Logistic regression produces binary output (yes or no). There is also a moving crossover average to show the past hours report. There was less accuracy in result of logistic regression when we shortened the data. Then we gave large dataset and the result was up to the mark. To conclude, we have completed our thesis work and we observed some curves and binary decisions in result. We are very much focused about the future plan as in ICU there are many machines but for resources, time and data limitations we had to work with only three machines. In future we have plans to automate the ICU unit totally. For that we need lots of resources and prepared dataset. We want to implement this system in real ICU. Then we will be able to train the system to make it even more accurate.

6. Reference:

1. Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S.: Big data: the future of biocuration. Nature 455(7209), 47–50 (2008)
2. Fieschi, M., Dufour, J.C., Staccini, P., Gouvernet, J., Bouhaddou, O.: Medical decision support systems: old dilemmas and new paradigms. Methods Inf. Med. 42(3), 190–198 (2003)
3. Bates, D.W., Kuperman, G.J., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M., Middleton, B.T.: Commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J. Am. Med. Inform. Assoc. 10(6), 523–530 (2003)
4. Wilk, S., Michalowski, W., O'Sullivan, D., Farion, K., Sayyad-Shirabad, J., Kuziemsky, C., Kukawka, B.: A task-based support architecture for developing point-of-care clinical decision support systems for the emergency department. Methods Inf. Med. 52(1), 18–32 (2013)
5. Lisboa, P.J., Taktak, A.F.G.: The use of artificial neural networks in decision support in cancer: a systematic review. Neural Netw. 19(4), 408–415 (2006)
6. Neill, D.B.: Using artificial intelligence to improve hospital inpatient care. IEEE Intell. Syst. 28(2), 92–95 (2013) 10.Baldi, P., Brunak, S.: Bioinformatics: The Machine Learning Approach. MIT Press, Cambridge (2001)
7. Min, S., Lee, B., Yoon, S.: Deep Learning in Bioinformatics. Brief. Bioinform., bbw068 (2016)
8. Libbrecht, M.W., Noble, W.S.: Machine learning applications in genetics and genomics. Nat. Rev. Genet. 16(6), 321–332 (2015)
9. Leung, M.K., Delong, A., Alipanahi, B., Frey, B.J.: Machine learning in genomic medicine: a review of computational problems and data sets. Proc. IEEE 104(1), 176–197 (2016)
10. Ben Wellner, Joan Grand, Elizabeth Canzone, Matt Coarr, Patrick W Brady, "Predicting Unplanned Transfers to the Intensive Care Unit: A Machine Learning Approach Leveraging Diverse Clinical Elements", JMIR Medical Informatics, vol. 5, pp. e45, 2017.
11. Johnson, Alistair Edward William et al. "Machine Learning and Decision Support in Critical Care." Proceedings of the IEEE (2016).
12. P. G. H. Metnitz et al., "SAPS 3—From evaluation of the patient to evaluation of

13. the intensive care unit. Part 1: Objectives, methods and cohort description," Intensive Care Med., vol. 31, no. 10, pp. 1336–1344, Oct. 2005.
14. A. Vesin et al., "Reporting and handling missing values in clinical studies in intensive care units," Intensive Care Med., vol. 39, no. 8, pp. 1396–1404, 2013.
15. R. Dybowski, P. Weller, R. Chang, and V. Gant, "Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm," Lancet, vol. 347, no. 9009, pp. 1146–1150,Apr. 1996.
16. L. S. Wong and J. D. Young, "A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks," Anaesthesia, vol. 54, no. 11, pp. 1048–1054, Nov. 1999.
17. G. Clermont, D. Angus, S. DiRusso, M. Griffin, and W. Linde-Zwirble, "Predicting hospital mortality for patients in the intensive care unit: A comparison of artificial neural networks with logistic regression models," Crit. Care Med., vol. 29, no. 2, pp. 291–296, 2001.
18. I. Silva, G. B. Moody, D. J. Scott, L. A. Celi, and R. G. Mark, "Predicting in-hospital mortality of ICU patients: The PhysioNet/ Computing in Cardiology Challenge 2012," Comput. Cardiol., vol. 39, pp. 245–248, 2012.
19. L. Lehman, T. Kyaw, G. Clifford, and R. Mark, "A temporal search engine for a massive multi-parameter clinical information database," in Proc. Comput. Cardiol., 2007, pp. 637–640.
20. N. Black, "High-quality clinical databases: Breaking down barriers," Lancet, vol.353, no. 9160, pp. 1205–1206, 1999.
21. M. Ghassemi et al., "A multivariate time series modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data," 2015.
22. L.-W. Lehman, M. Saeed, W. Long, J. Lee, and R. Mark, "Risk stratification of ICU patients using topic models inferred from unstructured progress notes," in Proc. AMIA Annu. Symp. Proc., 2012, vol. 2012, p. 505.
23. S. A. Collins et al., "Relationship between nursing documentation and patients' mortality," Amer. J. Crit. Care, vol. 22, no. 4, pp. 306–313, 2013.
24. Morris AH. Rational use of computerized protocols in the intensive care unit. Crit Care 2001 Oct;5(5):249-254.

25. L. H. Lehman et al., “A physiological time series dynamics-based approach to patient monitoring and outcome prediction,” IEEE J. Biomed. Health Inf., vol. 19, no. 3, pp. 1068–1076, 2015.
26. D. Clifton, D. Wong, L. Clifton, R. Pullinger, and L. Tarassenko, “A large-scale clinical validation of an integrated monitoring system in the emergency department,” IEEE Trans. Inf. Technol. Biomed., vol. 17, no. 4, pp. 835–877, 2013.
27. Vincent JL. Critical care: where have we been and where are we going? Crit Care 2013;17(1):S2
28. Harrison DA, Parry GJ, Carpenter JR, et al. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. Crit Care Med. 2007;35:1091–1098.
29. Render ML, Deddens J, Freybert R, et al. Veterans Affairs intensive care unit risk adjustment model: Validation, updating, recalibration. Crit Care Med. 2008;36:1031–1042.
30. Moody G, Mark R. A database to support development and evaluation of intelligent intensive care monitoring. Compute Cardio l. 1996;33:657–660.
31. Norris P, Dawant B. Closing the loop in ICU decision support: Physiologic event detection, alerts, and documentation. Proc AMIA Symp; 2001. pp. 498–502.
32. Wright A, Sittig DF, Ash JS, et al. Clinical decision support capabilities of commercially-available clinical information systems. J Am Med Inform Assoc. 2009;16:637–644.
33. Hug CW. PhD thesis. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science; 2009. Predicting the risk and trajectory of intensive care patients using survival models. Available at: http://dspace.mit.edu/handle/1721.1/33957.
34. R. Pirracchio et al., “Mortality prediction in intensive care units with the super ICU learner algorithm (sicula): A population-based study,” Lancet Respirat. Med., vol. 3, no. 1, pp. 42–52, 2015.
35. A. Smith, M. West, K. Gordon, M. Knapp, and I. Trimble, “Monitoring kidney transplant patients,” The Statistician, vol. 32, pp. 46–54, 1983.
36. G. Clifford, F. Azuaje, and P. McSharry, Advanced Methods and Tools for ECG Data Analysis. Boston, MA, USA: Artech House, 2006.

37. G. S. Power and D. A. Harrison, “Why try to predict ICU outcomes?” Current Opinion Crit. Care, vol. 20, no. 5, pp. 544–549, 2014.
38. S. Nizami, J. R. Green, and C. McGregor, “Implementation of artifact detection in critical care: A methodological review,” IEEE Rev. Biomed. Eng., vol. 6, pp. 127–142, 2013.
39. W. Zong, G. Moody, and R. Mark, “Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure,” Med. Biol. Eng. Comput., vol. 42, no. 5, pp. 698–706, 2004.
40. F. Plesinger, P. Klimes, J. Halamek, and P. Jurak, “False alarms in intensive care unit monitors : Detection of life-threatening arrhythmias using elementary algebra, descriptive statistics and fuzzy logic, ” in Proc. Comput. Cardiol. Conf . , 2015, pp. 1–4.
41. G. D. Clifford, W. J. Long, G. B. Moody, and P. Szolovits, “Robust parameter extraction for decision support using multimodal intensive care data,” Philosoph. Trans. A, Math. Phys. Eng. Sci., vol. 367, no. 1887, pp. 411–429, Jan. 2009.
42. C. W. Hug, G. D. Clifford, and A. T. Reisner, “Clinician blood pressure documentation of stable intensive care patients: An intelligent archiving agent has a higher association with future hypotension,” Crit. Care Med., vol. 39, no. 5, pp. 1006–1014, May 2011.
43. M. Saeed et al., “Multiparameter intelligent monitoring in intensive care (MIMIC II): A public-access intensive care unit database,” Crit. Care Med., vol. 39, no. 5, pp. 952–960, May 2011.
44. M. Saeed, C. Lieu, G. Raber, and R. G. Mark, “MIMIC II: A massive temporal ICU patient database to support research in intelligent patient monitoring,” Comput .Cardiol., vol. 29, pp. 641–644, 2002.
45. I. Neamatullah et al., “Automated de-identification of free-text medical records,” BMC Med. Inf. Decision Making, vol. 8, no. 1, p. 32, 2008.

46. M. C. Prosperi et al., “Investigation of expert rule bases, logistic regression, non-linear machine learning techniques for predicting response to antiretroviral treatment,” Antivir. Ther., vol. 14, no. 3, pp. 433–442, 2009.
47. C. Orphanidou et al., “Signal quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring,” IEEE J. Biomed. Health Inf., vol. 19, no. 3, pp. 832–838, 2015.
48. T. W. Liao, “Clustering of time series data—A survey,” Pattern Recognit., vol. 38, no. 11, pp. 1857–1874, 2005.
49. Johnson, A. E. W. et al. MIMIC-III, a freely accessible critical care database. Sci. Data 3:160035 doi: 10.1038/sdata.2016.35 (2016).

-------------------- ***The End*** --------------------