**IMPROVING PREDICTION PERFORMANCE OF DIFFERENT STAGES OF SEPSIS USING MACHINE LEARNING ALGORITHM**

**A PROJECT REPORT**

***Submitted by***

**R HARIKRISHNA 312316104055**

**JASON RAYEN 312316104063**

***in the partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**

**ST. JOSEPH’S COLLEGE OF ENGINEERING, CHENNAI**

**ANNA UNIVERSITY:: CHENNAI 600 025**

**MAY 2020**

**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report **“IMPROVING PREDICTION PERFORMANCE OF DIFFERENT STAGES OF SEPSIS USING MACHINE LEARNING ALGORITHM”**is the bonafide work of **R HARIKRISHNA (312316104055) and JASON RAYEN (312316104063)**who carried out the project work under my supervision.

|  |  |
| --- | --- |
| **SIGNATURE Dr. A. CHANDRASEKAR, M.E., Ph.D.,**  **HEAD OF THE DEPARTMENT**  Professor & Head  Department Of  Computer Science and Engineering  St. Joseph’s College of Engineering  Chennai-600119. | **SIGNATURE**  **Mrs. P.N. JEIPRATHA, M.E.,(Ph.D)**  **SUPERVISOR**  Assistant Professor Department Of  Computer Science and Engineering  St. Joseph’s College of Engineering  Chennai-600119. |

Submitted for the Viva-Voce held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

The contentment and elation that accompany the successful completion of any work would be incomplete without mentioning the people who made it possible.

Words are inadequate in offering our sincere thanks and performed gratitude to our beloved **Dr. B. Babu Manoharan, M.A., M.B.A., Ph.D.** Chairman and Managing Director, St. Joseph’s Group of Institutions for providing an opportunity to study in his esteemed Institution.

We also express our sincere thanks and gratitude to **Ms. B. Jessie Priya, M.Com.** and **Mr. B. Shashi Sekar, M.Sc.,** Directors, St. Joseph’s College of Engineering for extending their helping hands at all times during the course.

We are extremely happy and wish to express our sincere thanks and gratitude to our beloved **Dr. Vaddi Seshagiri Rao M.E., M.B.A., Ph.D.,** Principal, St. Joseph’s College of Engineering for his encouragement throughout the course.

We also wish to express our sincere thanks and most heartfelt sense of gratitude to **Dr. A. Chandrasekar, M.E., Ph.D.,** Head of the Department of Computer Science and Engineering, for his dedication, commendable support and encouragement for the completion of project work with perfection.

We convey our sincere thanks to our beloved project coordinator as well as our supervisor **Mrs. P.N. Jeipratha, M.E., (Ph.D).,** Assistant Professor, Department of Computer Science and Engineering, who was always a constant source of inspiration and encouragement to us during the course.

Last but not the least we thank our family members and friends who have been the greatest source of support for us.

**ABSTRACT**

Sepsis which known to be Septicemia, is alife-threatening condition that arises when the body's response to [infection](https://en.wikipedia.org/wiki/Infection" \o "Infection) causes injury to its tissues and organs,later on leads to death. Sepsis is an [inflammatory immune response](https://en.wikipedia.org/wiki/Immune_system" \l "Inflammation" \o "Immune system) triggered by an infection.Undetected sepsis can progress to severe sepsis and septic shock.The risk of death from sepsis is as high as 30%, as high as 50% from severe sepsis, and up to 80% from septic shock. Sepsis requires immediate treatment with [intravenous fluids](https://en.wikipedia.org/wiki/Intravenous_fluids" \o "Intravenous fluids) and [antimicrobials](https://en.wikipedia.org/wiki/Antimicrobial). Delay in predicting and identifying the various stages is the major concern. It may at least take two to three days to get the result. By the time the organs starts to dysfunction or tissue hypoperfusion. Children, particularly newborns and young infants, can be more susceptible in developing Sepsis.In existing system, the medical teams diagnose the disease from the patient’s medical history, symptoms, a physical exam and test. This approach promotes the early prediction of various sepsis level and could act as  prognostic stratification system .

**TABLE OF CONTENTS**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.NO** | **TITLE** | | | | | **PAGE NO**. |
|  | ABSTRACT | | | | | i |
|  | LIST OF FIGURES | | | | | viii |
|  | LIST OF ABBREVATIONS | | | | | ix |
| 1 | INTRODUCTION | | | | | 1 |
|  | 1.1 | GENERAL INTRODUCTION - SEPSIS | | | | 1 |
|  | 1.2 | DOMAIN OVERVIEW | | | | 2 |
|  |  | 1.2.1 | | | ARTIFICIAL INTELLIGENCE | 2 |
|  |  | 1.2.2 | | | UNDERSTANDING THE IMPORTANCE OF AI | 2 |
|  | 1.3 | FUTURE TECHNOLOGY OF INTELLIGENCE | | | | 3 |
|  |  | 1.3.1 | | | USE OF MULTIPLE TECHNOLOGIES IN ML | 4 |
|  |  |  | 1.3.1.1 | | PERSONALIZED COMPUTING ENVIRONMENT | 4 |
|  |  |  | 1.3.1.2 | | TUNED RECOMMENDATION ENGINE | 4 |
|  |  |  | 1.3.1.3 | | CLASSIFICATION AND REGRESSION FOR PREDICTIVE ANALYSIS | 4 |
|  |  |  | 1.3.1.4 | | CLUSTER ANALYSIS | 5 |
|  | 1.4 | MACHINE LEARNING - ELEMENTS | | | | 5 |
|  | 1.5 | CHARACTERSTICS OF MACHINE LEARNING | | | | 5 |
|  | 1.6 | ADVANTAGES OF MACHINE LEARNING | | | | 5 |
|  |  | 1.6.1 | | DENTIFYING TRENDS AND PATTERNS | | 5 |
|  |  | 1.6.2 | | LACK OF HUMAN INTERVENTION (AUTOMATION) | | 6 |
|  |  | 1.6.3 | | CONTINUOUS IMPROVEMENT | | 6 |
|  |  | 1.6.4 | | HANDLING MULTI-DIMENSIONAL AND MULTI-VARIENT DATA | | 6 |
|  |  | 1.6.5 | | WIDE APPLICATIONS | | 6 |
| 2 | LITERATURE SURVEY | | | | | 7 |
| 3 | SYSTEM CONCEPT | | | | |  |
|  | 3.1 | EXISTING SYSTEM | | | | 7 |
|  |  | 3.1.1 | | DISADVANTAGES | | 7 |
|  | 3.2 | PROPOSED SYSTEM | | | | 8 |
|  |  | 3.2.1 | | ADVANTAGES | | 8 |
| 4 | AIM AND SCOPE OF THE PRESENT INVESTIGATION | | | | | 9 |
|  | 4.1 | FUTURE SCOPE | | | | 9 |
|  | 4.2 | NEED OF THE PROJECT | | | | 9 |
|  |  | 4.2.1 | | PROBLEM DEFINITION | | 9 |
|  |  | 4.2.2 | | MODULES | | 9 |
|  |  |  | 4.2.2.1 | | DATA COLLECTION AND PREPROCESSING | 9 |
|  |  |  | 4.2.2.2 | | SEPSIS AND SEVERE SEPSIS MODEL | 10 |
|  |  |  | 4.2.2.3 | | SEPTIC SHOCK AND INTERFACE WITH UI | 10 |
|  |  | 4.2.3 | | WEB DEVELOPMENT MODULES | | 10 |
|  |  |  | 4.2.3.1 | | SEPSIS PREDICTION | 11 |
|  |  |  | 4.2.3.2 | | SEVERE SEPSIS PREDICTION | 11 |
|  |  |  | 4.2.3.3 | | SEPTIC SHOCK PREDICTION | 11 |
|  |  |  | 4.2.3.4 | | HOSPITAL SUGGESTION | 11 |
|  |  | 4.2.4 | | MODEL BUILDING | | 11 |
|  |  |  | 4.2.4.1 | | SUPERVISED LEARNING | 12 |
|  |  |  | 4.2.4.2 | | UNSUPERVISED LEARNING | 13 |
|  |  |  | 4.2.4.3 | | SEMI-SUPERVISED LEARNING | 13 |
|  |  |  | 4.2.4.4 | | REINFORCEMENT LEARNING | 14 |
| 5 | IMPLEMENTATION | | | | | 15 |
|  | 5.1 | METHODLOGIES | | | | 15 |
|  |  | 5.1.1 | | RANDOM FOREST | | 15 |
|  |  | 5.1.2 | | NEURAL NETWORK | | 15 |
|  | 5.2 | ACCURACY CALCULATION | | | | 16 |
|  |  | 5.2.1 | | CONFUSION MATRICES | | 16 |
|  |  | 5.2.2 | | FEASIBLITY STUDY | | 18 |
|  |  |  | 5.2.2.1 | | ECONIMICAL FEASIBILITY | 28 |
|  |  |  | 5.2.2.2 | | TECHINICAL FEASIBLITY | 28 |
|  |  |  | 5.2.2.3 | | SOCIAL FEASIBLITY | 28 |
|  |  | 5.2.3 | | REQUIREMENT SPECIFICATION | | 19 |
|  |  |  | 5.2.3.1 | | HARDWARE REQUIREMENTS |  |
|  |  |  | 5.2.3.2 | | SOFTWARE REQUIREMENTS |  |
| 6 | PERFORMANCE ANALYSIS | | | | | 20 |
|  | 6.1 | SYSTEM ARCHITECTURE | | | | 20 |
|  | 6.2 | PYTHON | | | | 20 |
|  | 6.3 | ANACONDA | | | | 21 |
|  | 6.4 | NUMPY | | | | 23 |
|  | 6.5 | PANDAS | | | | 24 |
|  | 6.6 | SEPSIS MODEL | | | | 27 |
|  | 6.7 | FLASK CODE | | | | 28 |
|  | 6.8 | UML DIAGRAMS | | | | 30 |
|  |  | 6.8.1 | | USECASE DIAGRAM FOR SEPSIS PREDICTION | | 31 |
|  |  | 6.8.2 | | CLASS DIAGRAM FOR SEPSIS PREDICTION | | 32 |
|  |  | 6.8.3 | | SEQUENCE DIAGRAM | | 32 |
|  |  | 6.8.4 | | ACTIVITY DIAGRAM | | 32 |
|  |  | 6.8.5 | | DATAFLOW DIAGRAM | | 34 |
| 7 | SYSTEM ANALYSIS | | | | | 35 |
|  | 7.1 | UNIT TESTING | | | | 35 |
|  | 7.2 | INTEGRATION TESTING | | | | 35 |
|  | 7.3 | FUNCTIONAL TEST | | | | 36 |
|  | 7.4 | WHITE BOX TESTING | | | | 36 |
|  | 7.5 | BLACK BOX TESTING | | | | 36 |
| 8 | SCREENSHOTS | | | | | 48 |
| 9 | CONCLUSIONS | | | | | 54 |
| 10 | REFERENCES | | | | | 55 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIG NO.** | **TITLE** | **PAGE NO.** |
| 3.2.1 | CLASSIFICATION | 12 |
| 3.2.2 | REGRESSION | 12 |
| 5.5.1 | SYSTEM ARCHITECTURE | 20 |

**LIST OF ABBREVATIONS**

|  |  |
| --- | --- |
| **ACRONYM** | **EXPANSION** |
| AI | ARTIFICIAL INTELLIGENCE |
| ML | MACHINE LEARNING |
| API | APPLICATION PROGRAM INTERFACE |

**CHAPTER I**

**INTRODUCTION**

**1.1 GENERAL INTRODUCTION - SEPSIS**

Sepsis indicated by national establishment of well being is characterized as an ailment where a body has a serious reaction to microbial diseases. Sepsis is a potentially life-threatening condition caused by the body's response to an infection. The body normally releases chemicals into the bloodstream to fight an infection. Sepsis occurs when cascading of such changes would damage multiple organ systems that leads to failure. Critically, there isn't one single microbes or infection that causes sepsis, rather any sort of pathogenic contamination can be the cause. This makes the treatment staggeringly troublesome, especially in territories where human services foundation isn't satisfactory enough to manage the different complexities. This has brought about sepsis harvesting a far higher loss of life in low to center salary nations. Various levels of sepsis is categorized as sepsis,severe sepsis, and septic shock.Sepsis seems to be influenced due to certain side effects like fever (more than 102'c) and heartbeat (more than 72 for every min) and breathing rate (more than 20 for every min). Severe sepsis leads to [poor organ function](https://en.wikipedia.org/wiki/Organ_dysfunction" \o "Organ dysfunction)ing or blood flow.(e.g. urinary tract contaminations, etc.). Septic shock causes low blood pressure and that cannot be improved after [fluid replacement](https://en.wikipedia.org/wiki/Fluid_replacement" \o "Fluid replacement).

The primary need of this prediction is to give early notice to patients and make alertness among the individuals about sepsis and its key elements. Machine learning concept is utilized and is termed to be the subset of computerized reasoning. ML is classified into several categories such as supervised learning, unsupervised learning,reinforcement learning and evolutionary learning. We proposed grouping calculation (Random Forest) so as to order the outcome and anticipate the sepsis of patients.

A boosting , a meta- algorithm which is a short form of adaptive random forest learning algorithm to improve performance. Arbitrary backwoods were proposed for the random woodland calculation and it can be utilized for both characterization and the relapse sort of issues. At the point of contrast with other grouping, relapse calculations in arbitrary woods gives more precision in light of fact that enormous choices of trees are estimated in irregular woodland.

**1.2 DOMAIN OVERVIEW**

**1.2.1 ARTIFICIAL INTELLIGENCE**

Artificial Intelligence (AI) ,termed as machine intelligence can be defined as the study of intelligent agents which are the device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals Colloquially. Artificial intelligence describes the machines (or systems) that mimic "cognitive" functions that humans associate with the [human mind](https://en.wikipedia.org/wiki/Human_mind" \o "Human mind), such as "learning" and "problem solving" . A more or less flexible or efficient approach can be considered depending on the requirements, which influences the behaviour of artificial intelligent.AI can be viewed in various perspectives. From the perspective of intelligence, artificial intelligence is making machines "intelligent" -- pretending to be how a human acts.The inability to distinguish the computer responses from human responses is termed as turing test. Expert problem solving define restricting the domain to allow significant relevant knowledge. From a business perspective, AI is a set of powerful tools that could utilize specific methodologies to solve business problems.From a programming perspective, it encomposes the study of symbolic programming, problem solving, and searching.Typically AI programs focuses on symbols rather than numeric processing. Problem solving mean to achieve goals where as search , the seldom that access the solution directly and could include a variety of techniques.The computer is interrogated by a human via a teletype.

**1.2.2 UNDERSTANDING THE IMPORTANCE OF AI**

* Game Playing
* Speech Recognition
* Natural Language
* Computer Vision
* Heuristic Classification
* Expert Systems

Machine learning is an application of artificial intelligence (AI) that provides the system,the ability to learn automatically and improve the performance from experience without being programmed explicitly. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The models are exposed to new data and are able to adapt independently.Learning achieved through previous computations to produce reliable, and optimized outputs.

* Supervised Learning
* Unsupervised Learning
* Reinforcement Learning

Deep learning which is known to be deep structured learning or differential programming is part of a broader family of [machine learning](https://en.wikipedia.org/wiki/Machine_learning" \o "Machine learning) methods based on [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_networks" \o "Artificial neural networks) with [representation learning](https://en.wikipedia.org/wiki/Representation_learning" \o "Representation learning). Deep learning architectures categorized as [deep neural networks](https://en.wikipedia.org/wiki/Deep_learning" \l "Deep_neural_networks), [deep belief networks](https://en.wikipedia.org/wiki/Deep_belief_network" \o "Deep belief network), [recurrent neural networks](https://en.wikipedia.org/wiki/Recurrent_neural_networks" \o "Recurrent neural networks) and [convolutional neural networks](https://en.wikipedia.org/wiki/Convolutional_neural_networks" \o "Convolutional neural networks) can be applied to various fields including [computer vision](https://en.wikipedia.org/wiki/Computer_vision" \o "Computer vision), [speech recognition](https://en.wikipedia.org/wiki/Automatic_speech_recognition" \o "Automatic speech recognition), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing" \o "Natural language processing), [audio recognition](https://en.wikipedia.org/wiki/Audio_recognition" \o "Audio recognition), social network filtering, [machine translation](https://en.wikipedia.org/wiki/Machine_translation" \o "Machine translation), [bioinformatics](https://en.wikipedia.org/wiki/Bioinformatics" \o "Bioinformatics), [drug design](https://en.wikipedia.org/wiki/Drug_design" \o "Drug design), medical image analysis, material inspection and [board game](https://en.wikipedia.org/wiki/Board_game" \o "Board game) programs, where they have produced results comparable to and in some cases surpassing human expert performance.

NLP(Natural Language Processing) is a field in machine learning with the ability of a computer to understand, analyze, manipulate, and potentially generate a human understandable language.The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable.Most NLP techniques rely on machine learning .

**1.3 FUTURE TECHNOLOGY OF INTELLIGENCE**

ML will be an integral part of AI systems that could be large or small.As ML assumes increased importance in business applications, there is a strong possibility of this technology being offered as a Cloud-based service known as Machine Learning-as-a-Service (MLaaS).Connected AI systems will enable ML algorithms to “continuously learn,” based on newly emerging information on the internet.There will be a big rush among hardware vendors to enhance CPU power to accommodate ML data processing. More accurately, hardware vendors will be pushed to redesign their machines to do justice to the powers of ML. Machine Learning will help machines to make better sense of context and meaning of data.

**1.3.1 USE OF MULTIPLE TECHNOLOGIES IN ML**

The emergence of IoT has benefited Machine Learning in many ways. The use of multiple technological strategies to achieve better learning is currently in practice in ML.In future more “collaborative learning” by utilizing multiple technologies is probable.

**1.3.1.1 PERSONALIZED COMPUTING ENVIRONMENT**

Developers will have access to API kits to design and deliver “more intelligent application.” In a specific way, this effort is akin to “assisted programming.” Through these API kits, developers will easily embed facial, speech, or vision-recognition features into their systems.Quantum Computing will greatly enhance the speed of execution of ML algorithms in high-dimensional vector processing. This will be the next conquest in the field of ML research.Future advancement in “unsupervised ML algorithms” will lead to higher business outcomes.

**1.3.1.2 TUNED RECOMMENDATION ENGINES**

ML-enabled services will become more accurate and relevant in future. For example, the recommendation Engines of the future will be far more relevant and closer to an individual user’s personal preferences and tastes.

**1.3.1.3 CLASSIFICATION AND REGRESSION FOR PREDICTIVE ANALYSIS** Classification is the process of finding a model that describes and distinguishes data classes or concepts. The model are derived based on the analysis of a set of training data . The model is used to predict the class label of objects for which the class label is unknown. There are many other methods for constructing classification models, such as naïve Bayesian classification, support vector machines, and k-nearest-neighbor classification.

**1.3.1.4 CLUSTER ANALYSIS**

Classification and Regression, which analyze class-labeled (training) data sets where as Clustering analyze the data objects without consulting class labels. Clustering can be used to generate class labels for a group of data. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity.

**1.4 MACHINE LEARNING - ELEMENTS**

* Data Models.
* Algorithms.
* Data sets.
* Features and Feature Extraction Techniques.
* Tunable Machine Learning and Advanced Threats.

**1.5 CHARACTERSTICS OF MACHINE LEARNING**

* The ability to perform automated data visualization
* Automation at its best
* Customer engagement like never before
* The ability to take efficiency to the next level when merged with IoT
* The ability to change the mortgage market
* Accurate data analysis
* Business intelligence at its best.

**1.6 ADVANTAGES OF MACHINE LEARNING**

#### **1.6.1 IDENTIFYING TRENDS AND PATTERNS**

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

#### **1.6.2. LACK OF HUMAN INTERVENTION (AUTOMATION)**

ML supports not to babysit the project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus software, they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

#### **1.6.3. CONTINUOUS IMPROVEMENT**

As ML algorithms  gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

#### **1.6.4. HANDLING MULTI-DIMENSIONAL AND MULTI-VARIENT DATA**

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

#### **1.6.5 WIDE APPLICATIONS**

#### Machine learning is a form of data analysis that automates analytical model building.Using algorithms that continuously access and learn from data, machine learning enables computers to access hidden insights. However, discoveries occur without programming systems to explicitly look for these digital treasures. Machine learning can easily consume unlimited amounts of data with timely analysis and assessment. Machine learning algorithms tend to operate at expedited levels.Applying machine learning to practical applications and scenarios is simply vital.

* Churn analysis
* Customer leads and conversion
* Customer defections

**CHAPTER 2**

**LITERATURE SURVEY**

**CLINICAL DATA ANALYSIS: AN OPPORTUNITY TO COMPARE MACHINE LEARNING METHODS**

**Author:** A Salcedo-Bernal, M. P. Villamil-Giraldo, and A. D. Moreno-Barbosa.

**ABSTRACT**

It has made use of multiple machine learning techniques that have been used successfully in clinical data analysis. However, there is little information about the parameter configurations, the required data transformations to prepare the data used to train and evaluate the models and the impact of these decisions in the accuracy of the predictive model. This research tackles these issues, using the clinical data of MIMICII to build features from physiological measure patterns to predict the decease of patients inside the hospital in the next 24 hours, building predictive models based on Logistic Regression, Neural Networks, Decision Trees and Nearest Neighbors. In particular, we use data associated to physiological measures of 3220 patients, where 2385 left the hospital alive and 835 passed in the hospital. The results show that the chosen strategy for building features from physiological data gives good results with Neural Networks and Logistic Regression with radial kernel models and the parameter configuration plays a fundamental role in the models performance.

**ANINTEGRATED DATA MINING APPROACH TO REAL-TIME CLINICAL MONITORING AND DETERIORATION WARNING**

**Author:** Y. Mao, W. Chen, Y. Chen, C. Lu, M. Kollef, and T. Bailey.

**ABSTRACT**

Clinical study found that early detection and intervention are essential for preventing clinical deterioration in patients, for patients both in intensive care units (ICU) as well as in general wards but under real-time data sensing (RDS). In this paper, we develop an integrated data mining approach to give early deterioration warnings for patients under real-time monitoring in ICU and RDS. Existing work on mining real-time clinical data often focus on certain single vital sign and specific disease. In this paper, we consider an integrated data mining approach for general sudden deterioration warning. We synthesize a large feature set that includes first and second order time-series features, detrended fluctuation analysis (DFA), spectral analysis, approximative entropy, and cross-signal features. We then systematically apply and evaluate a series of established data mining methods, including forward feature selection, linear and nonlinear classification algorithms, and exploratory under sampling for class imbalance. An extensive empirical study is conducted on real patient data collected between 2001 and 2008 from a variety of ICU. Results show the benefit of each of the proposed techniques, and the final integrated approach significantly improves the prediction quality. The proposed clinical warning system is currently under integration with the electronic medical record system at Barnes-Jewish Hospital in preparation for a clinical trial. This work represents a promising step toward general early clinical warning which has the potential to significantly improve the quality of patient care in hospitals.

**PHYSIOMARKERS IN REAL-TIME CONTINUOUS PHYSIOLOGICAL DATA STREAMS PREDICTS ADULT SEPSIS ONSET EARLIER THAN CLINICAL PRACTICE**

**Author:** F. Van Wyk, A. Khojandi, R.L. Davis, R. Kamaleswaran.

**ABSTRACT**

Neonatal Intensive Care Unit maintain and support life during the critical period of premature development. This research presents the challenges, trends and opportunities for integrated real time neonatal clinical decision support. We demonstrated this potential using environment known as Artemis, a clinical decision support system. A review of the current devices in the intensive care unit and neonatal practice shows the current environment and our perspective for the future of the neonatal clinical decision support. The study demonstrates that Artemis will be able to incorporate new data streams from infusion pumps, EEG monitors and cerebral oxygenation monitors innovating the practice and improving the clinical support.

**NATIONAL INPATIENT HOSPITAL COSTS: THE MOST EXPENSIVE CONDITIONS BY PAYER**

**Author:** C.M. Torio, and B.J Moore.

**ABSTRACT**

This Statistical Brief presents data from the Healthcare Cost and Utilization Project (HCUP) on costs of inpatient hospital stays in the United States in 2011. This report describes the distribution of costs by expected primary payer and illustrates the conditions accounting for the largest percentage of each payer’s hospital costs. The primary payers examined are Medicare, Medicaid, private insurance, and the uninsured. The hospital costs represent the hospital’s cost to produce the services—not the amount paid for services by payers—and they do not include the physician fees associated with the hospitalization. All differences between estimates noted in the text are statistically significant at the .05 level or better.

**EARLY SEPSIS DETECTION IN CRITICAL CARE PATIENTS USING MULTISCALE BLOOD PRESSURE AND HEART RATE DYNAMICS**

**Author:** S. P.Shashikumar, M.D. Stanley, I. Sadiq , Q. Li, A.Holder, G.D. Clifford, & S. Nemati.

**ABSTRACT**

Sepsis remains a leading cause of morbidity and mortality among intensive care unit (ICU) patients. For each hour treatment initiation is delayed after diagnosis, sepsis-related mortality increases by approximately 8%. Therefore, maximizing effective care requires early recognition and initiation of treatment protocols. Antecedent [signs and symptoms](https://www.sciencedirect.com/topics/medicine-and-dentistry/physical-disease-by-body-function" \o "Learn more about Physical Disease by Body Function from ScienceDirect's AI-generated Topic Pages) of sepsis can be subtle and unrecognizable (e.g., loss of autonomic regulation of vital signs), causing treatment delays and harm to the patient. In this work we investigated the utility of high-resolution blood pressure (BP) and heart rate (HR) [times series](https://www.sciencedirect.com/topics/medicine-and-dentistry/time-series-analysis" \o "Learn more about Time Series Analysis from ScienceDirect's AI-generated Topic Pages) dynamics for the early prediction of sepsis in patients from an urban, academic hospital, meeting the third international consensus definition of sepsis (sepsis-III) during their ICU admission. Using a multivariate modeling approach we found that HR and BP dynamics at multiple time-scales are independent predictors of sepsis, even after adjusting for commonly measured clinical values and patient demographics and comorbidities. Earlier recognition and diagnosis of sepsis has the potential to decrease sepsis-related morbidity and mortality through earlier initiation of treatment protocols.

**CASCADING CLASSIFIERS FOR NAMED ENTITY RECOGNITION IN CLINICAL NOTES**

**Author:** Y Wang, and J. Patrick.

**ABSTRACT**

Clinical named entities convey great deal of knowledge in clinical notes. This paper investigates named entity recognition from clinical notes using machine learning approaches. We present a cascading system that uses a Conditional Random Fields model, a Support Vector Machine and a Maximum Entropy to reclassify the identified entities in order to reduce misclassification. Voting strategy was employed to determine the class of the recognised entities between the three classifiers. The experiments were conducted on a corpus of 311 manually annotated admission summaries form an Intensive Care Unit. The recognition of 10 types of clinical named entities using 10 fold cross-validation achieved an overall results of 83.3 F-score. The reclassifier effectively increased the performance over stand-alone CRF models by 3.35 F-score.

**APOPTOSIS AND CASPASES REGULATE DEATH AND INFLAMMATION IN SEPSIS. NATURE REVIEWS IMMUNOLOGY**

**Author:** R.S. Hotchkiss And D.W Nicholson

**ABSTRACT**

Although the prevailing concept has been that mortality in sepsis results from an unbridled hyper-inflammatory cytokine-mediated response, the failure of more than 30 clinical trials to treat sepsis by controlling this cytokine response requires a 'rethink' of the molecular mechanism underpinning the development of sepsis. As we discuss here, remarkable new studies indicate that most deaths from sepsis are actually the result of a substantially impaired immune response that is due to extensive death of immune effector cells. Rectification of this apoptotic–inflammatory imbalance using modulators of caspases and other components of the cell-death pathway have shown striking efficacy in stringent animal models of sepsis, indicating an entirely novel path forward for the clinical treatment of human sepsis.

**REAL-TIME EARLY DETECTION OF ALLERGIC REACTIONS BASED ON HEART RATE VARIABILITY**

**Author:**R. Gutiérrez-Rivas

**ABSTRACT**

The popularisation of the concept of “Internet of Things” has promoted the fast increase of applications focused on obtaining information regarding people. For this reason, and thanks to the availability of the computing capacity of smart phones, over the last years a large number of low cost devices and applications have been marketed for analysing the health of users. In this thesis it is proposed to use ECG signals for early detection of allergic reactions. With this aim, a new QRS complex detection algorithm has been designed able to work in real time. This algorithm achieves an accuracy similar to those proposed by other authors, by reducing their computational complexity and the needed resources, which make it able to be implemented in portable platforms. In a previous study the effect that the occurrence of an allergic reaction causes in the heart rate variability was analysed, showing that it is noticeable even before the appearance of physical symptoms in most of the cases in which patients suffered an allergic reaction. However, the method proposed in this previous study is not suitable for detecting allergic reactions during real tests, since the computational complexity of the model designed requires hours of analysis to perform that detection. Moreover, the previous study only focused on food provocation tests in children under 12 years old. The study of the heart rate variability of allergic and non-allergic patients during provocation tests is continued in this work, with two main objectives: the designing of an algorithm capable of detecting allergic reactions in real time, and the extension of the study to include adults and drug provocation tests. The resulting algorithm has an accuracy similar to that proposed in the previous work and the achieved dose and length reduction of the provocation tests is similar as well. However, this algorithm is able to be implemented in a standalone portable device with limited resources and, what is more important, to perform the allergy reactions detection in realtime. Although the results are promising, this study should be interpreted as the beginning of further research, since it is necessary to spend more time and effort in acquiring new data to get a representative sample of the entire population of allergic patients in the case of both food and drug allergies.

**REAL-TIME PCR IN DETECTION AND QUANTITATION OF LEISHMANIA DONOVANI FOR THE DIAGNOSIS OF VISCERAL LEISHMANIASIS PATIENTS AND THE MONITORING OF THEIR RESPONSE TO TREATMENT**

**Author:** F. Hossain, P. Ghosh, M.A.A. Khan, M.S. Duthie, A.C. Vallur, A. Picone, R.F. Howard, S.G. Reed, and D. Mondal

**ABSTRACT**

Sustained elimination of Visceral Leishmaniasis (VL) requires the reduction and control of parasite reservoirs to minimize the transmission of *Leishmania donovani* infection. A simple, reproducible and definitive diagnostic procedure is therefore indispensable for the early and accurate detection of parasites in VL, Relapsed VL (RVL) and Post Kala-azar Dermal Leishmaniasis (PKDL) patients, all of whom are potential reservoirs of *Leishmania* parasites. To overcome the limitations of current diagnostic approaches, a novel quantitative real-time polymerase chain reaction (qPCR) method based on Taqman chemistry was devised for the detection and quantification of *L*. *donovani* in blood and skin. The diagnostic efficacy was evaluated using archived peripheral blood buffy coat DNA from 40 VL, 40 PKDL, 10 RVL, 20 cured VL, and 40 cured PKDL along with 10 tuberculosis (TB) cases and 80 healthy endemic controls. Results were compared to those obtained using a *Leishmania*-specific nested PCR (Ln-PCR). The real time PCR assay was 100% (95% CI, 91.19–100%) sensitive in detecting parasite genomes in VL and RVL samples and 85.0% (95% CI, 70.16–94.29%) sensitive for PKDL samples. In contrast, the sensitivity of Ln-PCR was 77.5% (95% CI, 61.55–89.16%) for VL samples, 100% (95%CI, 69.15–100%) for RVL samples, and 52.5% (95% CI, 36.13–68.49%) for PKDL samples. There was significant discordance between the two methods with the overall sensitivity of the qPCR assay being considerably higher than Ln-PCR. None of the assay detected *L*. *donovani* DNA in buffy coats from cured VL cases, and reduced infectious burdens were demonstrated in cured PKDL cases who remained positive in 7.5% (3/40) and 2.5% (1/40) cases by real-time PCR and Ln-PCR, respectively. Both assays were 100% (95% CI, 95.98–100) specific with no positive signals in either endemic healthy control or TB samples. The real time PCR assay we developed offers a molecular tool for accurate detection of circulating *L*. *donovani* parasites in VL, PKDL and RVL patients, as well as being capable of assessing response to treatment. As such, this real time PCR assay represents an important contribution in efforts to eliminate VL.

**ESTIMATINGCLASSIFICATIONERRORRATE:REPEATEDCROSS-VALIDATION, REPEATEDHOLD-OUTANDBOOTSTRAP**

**Author:** J. H. Kim.

**ABSTRACT**

We consider the accuracy estimation of a classifier constructed on a given training sample. The naive resubstitution estimate is known to have a downward bias problem. The traditional approach to tackling this bias problem is cross-validation. The bootstrap is another way to bring down the high variability of cross-validation. But a direct comparison of the two estimators, cross-validation and bootstrap, is not fair because the latter estimator requires much heavier computation. We performed an empirical study to compare the .632+ bootstrap estimator with the repeated 10-fold cross-validation and the repeated one-third holdout estimator. All the estimators were set to require about the same amount of computation. In the simulation study, the repeated 10-fold cross-validation estimator was found to have better performance than the .632+ bootstrap estimator when the classifier is highly adaptive to the training sample. We have also found that the .632+ bootstrap estimator suffers from a bias problem for large samples as well as for small samples.

**ECISION THRESHOLD ADJUSTMENT IN CLASS PREDICTION**

**Author:**J.J. Chen, C.A. Tsai, H. Moon, H. Ahn, J.J Young And C.H. Chen.

**ABSTRACT**

Standard classification algorithms are generally designed to maximize the number of correct predictions (concordance). The criterion of maximizing the concordance may not be appropriate in certain applications. In practice, some applications may emphasize high sensitivity (e.g., clinical diagnostic tests) and others may emphasize high specificity (e.g., epidemiology screening studies). This paper considers effects of the decision threshold on sensitivity, specificity, and concordance for four classification methods: logistic regression, classification tree, Fisher's linear discriminant analysis, and a weighted k-nearest neighbor. We investigated the use of decision threshold adjustment to improve performance of either sensitivity or specificity of a classifier under specific conditions. We conducted a Monte Carlo simulation showing that as the decision threshold increases, the sensitivity decreases and the specificity increases; but, the concordance values in an interval around the maximum concordance are similar. For specified sensitivity and specificity levels, an

optimal decision threshold might be determined in an interval around the maximum concordance that meets the specified requirement. Three example data sets were analyzed for illustrations.

**PREDICTING CLINICAL EVENTS VIA RECURRENT NEURAL NETWORKS**

**Author:**Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun.

**ABSTRACT**

Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another without losing substantial accuracy.

**COMPARATIVE ANALYSIS OF RECURRENT AND FINITE IMPULSE RESPONSE NEURAL NETWORKS IN TIME SERIES PREDICTION**

**Author:** M. Miljanovic.

**ABSTRACT**

The purpose of this paper is to perform evaluation of two different neural network architectures used for solving temporal problems, i.e. time series prediction. The data sets in this project include Mackey-Glass, Sunspots, and Standard & Poor's 500, the stock market index. The study also presents a comparison study on the two networks and their performance.

**AUTOMATED MASS DETECTION IN MAMMOGRAMS USING CASCADED DEEP LEARNING AND RANDOM FORESTS**

**Author:** N. Dhungel, G. Carneiro, and A. P. Bradley.

**ABSTRACT**

Mass detection from mammograms plays a crucial role as a pre- processing stage for mass segmentation and classification. The detection of masses from mammograms is considered to be a challenging problem due to their large variation in shape, size, boundary and texture and also because of their low signal to noise ratio compared to the surrounding breast tissue. In this paper, we present a novel approach for detecting masses in mammograms using a cascade of deep learning and random forest classifiers. The first stage classifier consists of a multi-scale deep belief network that selects suspicious regions to be further processed by a two-level cascade of deep convolutional neural networks. The regions that survive this deep learning analysis are then processed by a two-level cascade of random forest classifiers that use morphological and texture features extracted from regions selected along the cascade. Finally, regions that survive the cascade of random forest classifiers are combined using connected component analysis to produce state-of-the-art results. We also show that the proposed cascade of deep learning and random forest classifiers are effective in the reduction of false positive regions, while maintaining a high true positive detection rate. We tested our mass detection system on two publicly available datasets: DDSM-BCRP and INbreast. The final mass detection produced by our approach achieves the best results on these publicly available datasets with a true positive rate of 0.96 ± 0.03 at 1.2 false positive per image on INbreast and true positive rate of 0.75 at 4.8 false positive per image on DDSM-BCRP.

**AUTOMATIC IDENTIFICATION OF BLOOD VESSEL CROSS-SECTION FOR CENTRAL VENOUS CATHETER PLACEMENT USING A CASCADING CLASSIFIER**

**Author:** M. Ikhsan, K.K. Tan, A.S. Putra, T.H.S Chew, and C.F. Kong.

**ABSTRACT**

This paper presents a system that is able to automatically identify, segment, and track the cross-section of the internal jugular vein (IJV) and the common carotid artery (CCA) in an ultrasound image feed during a central venous catheter (CVC) placement procedure. The goal is to provide assistance to the practitioner in order to decrease the probability of complications stemming from inadvertent punctures of the CCA during the procedure. In the system, a modified Star algorithm is implemented to segment and track the blood vessel throughout an ultrasound video feed. A novel algorithm based on a cascading classifier is used to identify the location of the IJV and the CCA for two main tasks: (1) selecting the initial seed point at the start of tracking and (2) validating the segmentation results at each subsequent frame. The classifier uses shape features (vessel area, ellipse fitting error, vessel depth, vessel eccentricity) and pixel-based features (pixel intensity and the histogram of oriented gradients descriptor) to differentiate between vessel and non-vessel structures and also differentiate between the IJV and the CCA. Evaluated on a database of 800 ultrasound images containing the cross-section of both vessels, the cascading classifier was able to identify the IJV and the CCA in 92.25% and 85.13% of the images respectively without anyinitialization from the user at a maximum processing rate of 40.65 frames per second. This allows identification to be conducted in real-time with existing ultrasound machines.

**RAPIDOBJECTDETECTIONUSINGABOOSTEDCASCADEOF SIMPLE FEATURES**

**Author:** P. Viola, and M. Jones

**ABSTRACT**

This paper describes a machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. This work is distinguished by three key contributions. The first is the introduction of a new image representation called the "integral image" which allows the features used by our detector to be computed very quickly. The second is a learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers. The third contribution is a method for combining increasingly more complex classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The cascade can be viewed as an object specific focus-of-attention mechanism which unlike previous approaches provides statistical guarantees that discarded regions are unlikely to contain the object of interest. In the domain of face detection the system yields detection rates comparable to the best previous systems. Used in real-time applications, the detector runs at 15 frames per second without resorting to image differencing or skin color detection.

**CHAPTER 3**

**SYSTEM CONCEPT**

**3.1 EXISTING SYSTEM**

Physiological data streams of all patients are collected using sliding time windows. That is, at each time interval (e.g., each minute), as new observations are collected at the bedside, the time window moves forward and the earliest observations are discarded from the time windows to maintain their lengths.

Features are extracted from the data collected by each of these time windows. These features are used in real- time as an input to the model in Layer 1 to continuously predict the likelihood of developing sepsis for patients. At this point, using a hard threshold on the model-estimated probabilities, allows for predicting sepsis. For instance, for the patient .

Features are extracted from the probability streams collected by each of the secondary time windows. These features are used in real-time as an input to the model in Layer 2 to continuously predict the likelihood of developing sepsis for patients. This online approach to feature extraction highlights the novelty in our approach, apart from existing cascading classifiers. At this point, we use a hard threshold of 0.5 to determine if a patient has sepsis.

Random forest (RF) is a supervised learning algorithm that uses an ensemble of decision trees to build a ‘forest.’ An aggregated majority vote across trees can be used for classification . Contrary to conventional decision trees, the algorithm uses a random subset of features when splitting decision nodes which result, in general, in better and more stable performance compared to decision tree learning and bagging techniques.

**3.1.1 DISADVANTAGES**

* Less accuracy compare with proposed model
* Able to run only in python idle
* Uses single trained model

**3.2 PROPOSED SYSTEM**

Sepsis is a possibly dangerous condition brought about by the body reaction to a disease. The body regularly discharges synthetic concoctions into the circulatory system to battle a disease. Sepsis happens when the body reaction to these synthetic concoctions is out of equalization, activating changes that can harm numerous organ frameworks. Distinguishing Sepsis takes additional time before identifying the disease.It leads the patient to the next stage. In the existing system, there are numerous models that work for prediction but the forecast precision of the models are not effective and not satisfactory. We build an artificially intelligent way to deal with inconsistent break down and high-recurrence information. We show the abilities of this methodology for early recognition of patients in danger of sepsis, a conceivably dangerous inconvenience of a contamination, using high-frequency  physiological information .We made the different models for sepsis and its stages and the dataset comprises of almost 25000 samples with the goal that precision and expectation is progressively precise. The algorithms like Random Forest, Neural systems , SVM and Naive Bayes can be utilized for totrain the model. Stages like Sepsis, Severe Sepsis and Septic shock can be anticipated utilizing the separate AI calculation.

**3.2.1 ADVANTAGES**

* Increasingly Accurate Dataset and Model.
* Separate Model for each stage. Initial stage and final stage
* Interactive Model and Interfaced with Web Application.
* Able to run in mobile.

**CHAPTER 4**

**AIM AND SCOPE OF THE PRESENT INVESTIGATION**

4.**1 FUTURE SCOPE**

New age AI design structure is fused into non-uniform spectrogram estimation. An excessively human AI expectation machine will utilize every one of those Sepsis/Non-sepsis convictions in assessing how well an IV anti-toxin and an additional blood culture treatment will serve as Sepsis/Non-Sepsis utility capacities. This will move the relative need to either Sepsis or Non-sepsis normal utilities after sometime with a flipping Boolean irregular variable, Sepsis Label. We will at that point have the option to anticipate the beginning time of sepsis and relative needs of the mutual control of two convictions. Sepsis and non-sepsis convictions increment the degree on which the machine's choices will influence the utilities. Moreover, the beginning time of sepsis will be found out by the POMDP instead of determining expressively.

**4.2 NEED OF THE PROJECT**

**4.2.1 PROBLEM DEFINITION**

The fundamental point of this examination is to foresee whether the patient or an individual possesses sepsis or not. Here we actualize the preparation database of 10,000 subjects from Clinical Data from different data sources. Each line of the table furnishes an assortment of estimations simultaneously (e.g., pulse and oxygen level simultaneously). In view of this element we will anticipate the sepsis utilizing subset of AI administered learning module with precision score of 95%.

**4.2.2 MODULES**

* Data Collection and Preprocessing
* Creating Sepsis and Severe Sepsis Model
* Creating Septic shock and Interface with UI

**4.2.2.1. DATA COLLECTION AND PREPROCESSING**

Data collection is the process of collecting details about the sepsis like symptoms and the rate of possible counts(heart rate, respiratory rate,bilurumin count etc.,).Aftercollecting the data it has to be categorizedbased on the symptoms. Three datasets for the three models is preferred for efficient learning. It increases the accuracy rate as well as the perfectness of prediction model. Preprocessing is the process of preparing the data that can be understandable by the machine. After the selection of the features ,the data is used to train the model.

**4.2.2.2 SEPSIS AND SEVERE SEPSIS MODEL**

After the preprocessing of the data it can be used for training. We split the dataset as features and label and then it could be used for training the model. Sklearn is the package for implementing the algorithm.Different algorithms like random forest, SVM ,naive bayes and many can be used. And the algorithm could be based on their greater accuracy . The stages of sepsis are sepsis prediction , severe sepsis and septic shock.

**4.2.2.3 SEPTIC SHOCK AND INTERFACE WITH UI**

After the Sepsis and Severe sepsis a model for Septic shock has to be created . Now split the data set into train and test data. Apply training set for training the machine and test for prediction. After the prediction the accuracy for every algorithm is estimated and could finalize the perfect model.We need a web application to see the prediction results. But the machine learning and Web development belongs to different domain. Create the pipeline for implementing the machine learning model using the pickle package. The pickle package will store the machine learning model in various stages of prediction. After that user can give input and can get the output results.

**4.2.3 WEB DEVELOPMENT MODULES**

Flask is a web framework for Python , it provides functionality for building web applications, including managing HTTP requests and rendering templates. Flask application is used to create our API. Flask’s framework is more explicit and is easier to learn because it have less base code to implement a simple web-Application. In proposed system flask framework plays a vital role in predicting sepsis and its various stages.

**4.2.3.1. SEPSIS PREDICTION**

In this module, the user can give the necessary inputs like location, Respiratory rate, consciousness , Heratbeat rate , WBC count ,Temperature of the body , CRP range in order to predict whether the person have sepsis or not.

**4.2.3.2. SEVERE SEPSIS PREDICTION**

In this module, the severe sepsis can be detected if the result of sepsis prediction become positive the user can give the necessary inputs like urine –output, SPO2 in order to predict whether the person/ patient having severe sepsis or not.

**4.2.3.3 SEPTIC SHOCK PREDICTION**

In this module, the septic shock is detected if severe sepsis prediction seems to be positive. Septic shock is predicted based on the inputs of systolic/diastolic pressure, glucose , creatine , bilirubin, INR, platelets .

**4.2.3.4 HOSPITAL SUGGESTION**

Above predictions can show whether the person has sepsis according to its severity category. This module will the suggest the hospitals.

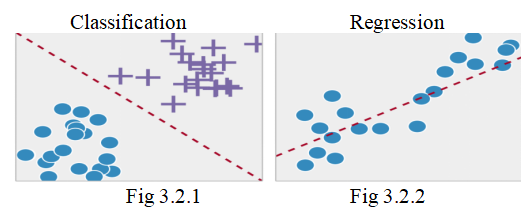
**4.2.4 MODEL BUILDING**

There is a wide range of machine learning algorithms to choose from, most of which are available in the python library [Scikit-learn](http://scikit-learn.org/stable/). However, most of the implementations of these algorithms do not accept sparse matrices as inputs, and since a large number of nominal features are chosen from n-gram features. It is imperative that we encode our features in a sparse matrix. Out of the algorithms that support sparse matrices in Scikit-learn, I ended up trying random forest with different input dataset.Best result on cross validating was obtained using random forest. The metricused here to guide the cross-validation is  [F-score](http://en.wikipedia.org/wiki/Precision_and_recall). Its a widely accepted metric when several samples are chosen from one category rather from different categories. In proposed n number of inputs to predict the various stages of sepsis. If the above metric is used to analyze the accuracy, which is estimated by the ratio of number of correct predictions to the total number of sepsis inputs, then a simple classifier which always predicts the sepsis stages and prediction would get a 99% accuracy. This is obviously very misleading so we need a better metric. We can do better by considering precision and recall for the various stages of sepsis. Precision is the number of inputs correctly identified divided by the total number of sepsis inputs classified as sepsis, while recall is the number of sepsis correctly identified divided by the total number of sepsis predicted in the cross validation set. Both precision and recall would be equal to 0% with a dumb classifier which always predicts load to be accepted, so these are already much better scores to quantify the quality of a sepsis model classifier. The F-score is simply the harmonic mean of precision and recall.Model Selection is done by the following types of machine learning:

* Supervised learning
* Unsupervised learning
* Semi-supervised learning
* Reinforcement learning

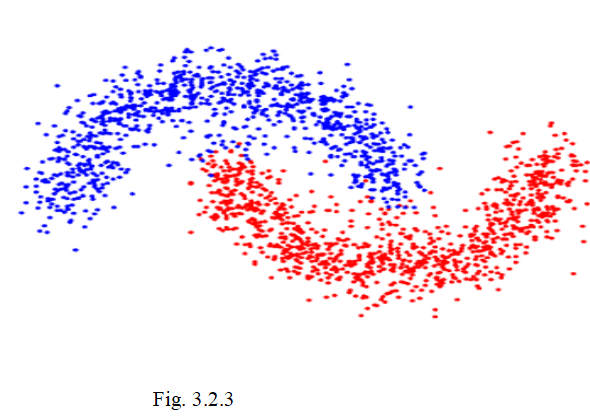
**4.2.4.1 SUPERVISED LEARNING**

Supervised learning is the task of inferring a function from labeled training data. By fitting to the labeled training set, we want to find the most optimal model parameters to predict unknown labels on other objects (test set). If the label is a real number, we call the task regression. If the label is from the limited number of values, where these values are unordered, then it is called as classification.



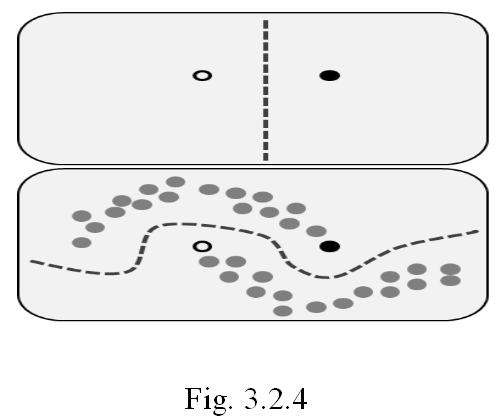
**4.2.4.2 UNSUPERVISED LEARNING**

In unsupervised learning we have less information about objects, in particular, the training set is unlabeled. It’s possible to observe some similarities between groups of objects and include them in appropriate clusters. Some objects can differ hugely from all clusters, hence these objects are assumed to be anomalies.



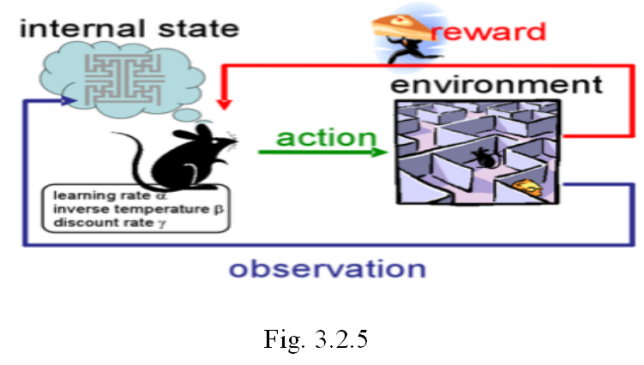
**4.2.4.3 SEMI-SUPERVISED LEARNING**

Semi-supervised learning tasks include both problems we described earlier: they use labeled and unlabeled data. That is a great opportunity for those who can’t afford labeling their data. The method allows us to significantly improve accuracy, because we can use unlabeled data in the train set with a small amount of labeled data .



**4.2.4.4 REINFORCEMENT LEARNING**

Reinforcement learning is not like any of the previous tasks because there are no labeled or unlabeled datasets here. RL is an area of machine learning concerned with how software agents ought to take actions in some environment to maximize the notion of cumulative reward.



If there is a robot in some strange place, it can perform the activities and get rewards from the environment. After each action the behavior gets more complex and clever, so the robot is training to behave in the most effective way for each step. Biologically, this is called adaptation to the natural environment.

**CHAPTER 5**

**IMPLEMENTATION**

**5.1 METHODOLOGIES**

**5.1.1 Random Forest**

Random forests or random decision forests are an [ensemble learning](https://en.wikipedia.org/wiki/Ensemble_learning) method for [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and other tasks that operates by constructing a multitude of [decision trees](https://en.wikipedia.org/wiki/Decision_tree_learning) at training time and outputting the class that is the [mode](https://en.wikipedia.org/wiki/Mode_(statistics)) of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of [overfitting](https://en.wikipedia.org/wiki/Overfitting) to their [training set](https://en.wikipedia.org/wiki/Test_set). The first algorithm for random decision forests was created by [Tin Kam Ho](https://en.wikipedia.org/wiki/Tin_Kam_Ho) using the [random subspace method](https://en.wikipedia.org/wiki/Random_subspace_method), which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.An extension of the algorithm was developed by [Leo Breiman](https://en.wikipedia.org/wiki/Leo_Breiman) and Adele Cutler, who registered "Random Forests" as a [trademark](https://en.wikipedia.org/wiki/Trademark) (as of 2019, owned by [Minitab, Inc.](https://en.wikipedia.org/wiki/Minitab)). The extension combines Breiman's "[bagging](https://en.wikipedia.org/wiki/Bootstrap_aggregating)" idea and random selection of features, introduced first by Ho and later independently by Amit and [Geman](https://en.wikipedia.org/wiki/Donald_Geman) in order to construct a collection of decision trees with controlled variance.

**5.1.2 NEURAL NETWORK**

A neural network is a [network or circuit](https://en.wikipedia.org/wiki/Biological_neural_network) of [neurons](https://en.wikipedia.org/wiki/Neuron), or in a modern sense, an [artificial neural network](https://en.wikipedia.org/wiki/Artificial_neural_network), composed of [artificial neurons](https://en.wikipedia.org/wiki/Artificial_neuron) or nodes.[[1]](https://en.wikipedia.org/wiki/Neural_network" \l "cite_note-1) Thus a neural network is either a [biological neural network](https://en.wikipedia.org/wiki/Biological_neural_network), made up of real biological neurons, or an artificial neural network, for solving [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) (AI) problems. The connections of the biological neuron are modeled as weights. A positive weight reflects an exciting connection, while negative values mean inhibitory connections. All inputs are modified by a weight and summed. This activity is referred to as a linear combination. Finally, an activation function controls the [amplitude](https://en.wikipedia.org/wiki/Amplitude) of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be −1 and 1.

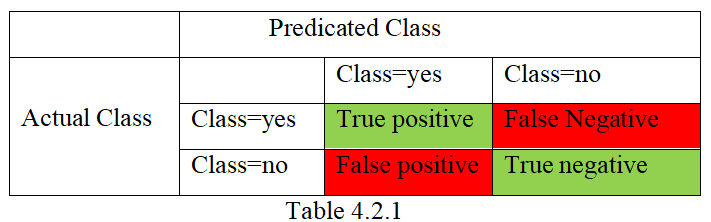
Unlike [von Neumann model](https://en.wikipedia.org/wiki/Von_Neumann_model) computations, artificial neural networks do not separate memory and processing and operate via the flow of signals through the net connections, somewhat akin to biological networks.

These artificial networks may be used for [predictive modeling](https://en.wikipedia.org/wiki/Predictive_modeling), adaptive control and applications where they can be trained via a dataset. Self-learning resulting from experience can occur within networks, which can derive conclusions from a complex and seemingly unrelated set of information.

**5.2 ACCURACY CALCULATION**

**5.2.1 CONFUSION MATRICES**

ROC curve is determined to be effective by looking at AUC (Area Under the Curve) and other parameters which are also called as Confusion Matrices. A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. All the measures except AUC can be calculated by using left most four parameters.



True positive and true negatives are the observations that are correctly predicted and are shown in green. We want to minimize false positives and false negatives so they are shown in red color. These terms are a bit confusing.

True Positives (TP) - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that the patient survived and predicted class tells you the same thing.

True Negatives (TN) - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this patient did not survive and predicted class tells you the same thing.

False Positives (FP) – When actual class is no and predicted class is yes. E.g. if actual class says this patient did not survive but predicted class tells you that this patient will survive.

False Negatives (FN) – When actual class is yes but predicted class in no. E.g. if actual class value indicates that this patient survived and predicted class tells you that patient will die.

Thus Accuracy, Precision, Recall and F1 score can be calculated.

Accuracy - Accuracy is the most intuitive performance measure and is merely a magnitude relation of properly foretold observation. One might imagine that, if we've high accuracy then our model is best. Yes, accuracy may be a reliable measure however you only have radially symmetrical datasets wherever values of false positive and false negatives area unit are virtually the same. Therefore, you have to appear at different parameters to judge the performance of your model. In the proposed system, the accuracy obtained is approximately 90%.

Accuracy = TP+TN/TP+FP+FN+TN

Precision - precision is that the magnitude relation of properly foretold positive observations to the entire foretold positive observations. High preciseness relates to the low false positive rate. 0.788 preciseness is attained.

Precision = TP/TP+FP

Recall (Sensitivity) - Recall is that the magnitude relation of properly foretold positive observations to the all observations in actual category - affirmative. The question recall answers is: Of all the passengers that actually survived, what percentage did we tend to label? We have got recall of 0.631 that is nice for this model as it’s on top of 0.5.

Recall = TP/TP+FN

F1 score - F1 Score is that the weighted average of preciseness and Recall. Therefore, this score takes each false positives and false negatives under consideration. Intuitively it's not as straightforward to know as accuracy, however F1 is sometimes a lot of helpful than accuracy, particularly if you have got AN uneven category distribution. Accuracy works best if false positives and false negatives have similar price. If the value of false positives and false negatives area unit terribly totally different, it’s higher to appear at each preciseness and Recall. In our case, F1 score is 0.701.

F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)

# **5.2.2 FEASIBLITY STUDY**

The objective of feasibility study is not only to solve the difficulty but also to obtain a sense of its scope. During the study, the problem definition was crystallized and aspects of the problem to be included in the system are determined. Consequently benefits are estimated with greater accuracy at this stage. The key considerations are:

* Economic feasibility
* Technical feasibility
* Social feasibility

**5.2.2.1 ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on an organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

**5.2.2.2 TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**5.2.2.3 SOCIAL FEASIBILITY**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system

**5.2.3 REQUIREMENT SPECIFICATION**

**5.2.3.1 HARDWARE REQUIREMENTS**

Processor : Pentium Dual Core 2.3 GHz

Hard Disk : 250 GB or Higher

Ram : 2 GB (Min)

**5.2.3.2 SOFTWARE REQUIREMENTS**

Operating System : Windows 7 or Higher

Languages used : Python(Pandas,Numpy,Sklearn)

Tools : Anaconda, Jupyter Notebook,Spyder

Backend : Machine Learning, Sqlite3.

**CHAPTER 6**

**PERFORMANCE ANALYSIS**

**6.1 SYSTEM ARCHITECTURE**

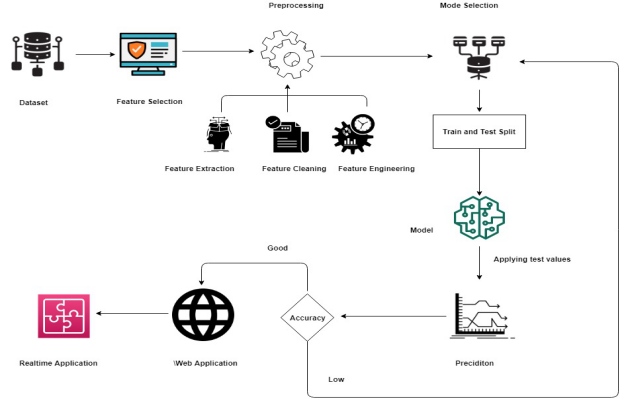


Fig. 5.1.1

**6.2 PYTHON**

Python is an interpreter, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed .Often, programmers fall in love with Python because of the increased productivity and flexibility it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.Python is a dynamic programming language which supports several different programming paradigms:

* Procedural programming
* Object oriented programming
* Functional programming

Standard: Python byte code is executed in the Python interpreter (similar to Java) → platform independent code

* Extremely versatile language - Website development, data analysis, server maintenance, numerical analysis.
* Syntax is clear, easy to read and learn (almost pseudo code)
* Common language
* Intuitive object oriented programming
* Full modularity, hierarchical packages
* Comprehensive standard library for many tasks
* Big community
* Simply extendable via C/C++, wrapping of C/C++ libraries
* Focus: Programming speed

6**.3 ANACONDA**

Anaconda is a [free and open-source](https://en.wikipedia.org/wiki/Free_and_open-source)[[5]](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)" \l "cite_note-5) distribution of the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) and [R](https://en.wikipedia.org/wiki/R_(programming_language)) programming languages for [scientific computing](https://en.wikipedia.org/wiki/Scientific_computing) ([data science](https://en.wikipedia.org/wiki/Data_science), [machine learning](https://en.wikipedia.org/wiki/Machine_learning) applications, large-scale data processing, [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics), etc.), that aims to simplify [package management](https://en.wikipedia.org/wiki/Package_management) and deployment. Package versions are managed by the [package management system](https://en.wikipedia.org/wiki/Package_manager) [conda](https://en.wikipedia.org/wiki/Conda_(package_manager)) The Anaconda distribution is used by over 12 million users and includes more than 1400 popular data-science packages suitable for Windows, Linux, and MacOS.

Anaconda enable the user to create virtual environments and install packages needed for data science and deep learning. With virtual environments specific package versions for a particular project or a tutorial without worrying about version conflicts can be installed.

Download Anaconda for your platform and choose the Python 3.6 version: [https://www.anaconda.com/download](https://www.anaconda.com/download/" \l "macos).

Anaconda, provides spider ,  conda, Python, Jupyter Notebook and hundreds of other open source packages.Conda is a package manager to manage virtual environment and install packages. some helpful commands using conda are:

#update conda in your default environment   
$ conda upgrade conda  
$ conda upgrade --all

# create a new environment with conda  
$ conda create -n [my-env-name]  
$ conda cerate -n [my-env-name] python=[python-version]

# activate the environment you created  
$ source activate [my-env-name]

# take a look at the environment you created  
$ conda info  
$ conda list

# install a package with conda and verify it's installed  
$ conda install numpy  
$ conda list

# take a look at the list of environments you currently have  
$ conda info -e

# remove an environment  
$conda env remove --name [my-env-name]

Conda vs Pip install

You can use either conda or pip for installation in an virtual environment created with conda. They are both open source package managers. Here are some differences:

* conda install — installs any software package.
* pip install — installs python packages only and it’s the defacto python package manager.

**6.4 NUMPY**

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more. At the core of the NumPy package, is the ndarray object. This encapsulates n-dimensional arrays of homogeneous data types, with many operations being performed in compiled code for performance. There are several important differences between NumPy arrays and the standard Python sequences:

* NumPy arrays have a fixed size at creation, unlike Python lists (which can grow dynamically). Changing the size of an array will create a new array and delete the original.
* The elements in a NumPy array are all required to be of the same data type, and thus will be the same size in memory. The exception: one can have arrays of (Python, including NumPy) objects, thereby allowing for arrays of different sized elements.
* NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data. Typically, such operations are executed more efficiently and with less code than is possible using Python’s built-in sequences.
* A growing plethora of scientific and mathematical Python-based packages are using NumPy arrays; though these typically support Python-sequence input, they convert such input to NumPy arrays prior to processing, and they often output NumPy arrays. In other words, in order to efficiently use much (perhaps even most) of today’s scientific/mathematical Python-based software, just knowing how to use Python’s built-in sequence types is insufficient - one also needs to know how to use NumPy arrays.

6**.5 PANDAS**

Data processing is important part of analyzing the data, because data is not all the time accessible in preferred format. Various dispensation are necessary before analyzing the data such as cleaning, restructuring or merging etc. Numpy, Scipy, Cython and Panda are the tools available in python which can be used fast processing of the data. Further, Pandas are built on the top of Numpy. Pandas provides rich set of functions to process various types of data. Further, working with Panda is fast, easy and more expressive than other tools. Pandas provides fast data processing as Numpy along with flexible data manipulation techniques as spreadsheets and relational databases. Lastly, pandas integrates well with matplotlib library, which makes it very handy tool for analyzing the data.

Pandas provides two very useful data structures to process the data i.e. Series and Data Frame . The Series is a one-dimensional array that can store various data types, including mix data types. The row labels in a Series are called the index. Any list, tuple and dictionary can be converted in to Series using ‘series’ .

Data Frame is the widely used data structure of pandas. Note that, Series are used to work with one dimensional array, whereas Data Frame can be used with two dimensional arrays. Data Frame has two different index i.e. column-index and row-index. The most common way to create a Data Frame is by using the dictionary of equal-length list as shown below. Further, all the spreadsheets and text files are read as Data Frame, therefore it is very important data structure of pandas.

Code :ml

**Code :predicition model**

# Import packages

import os

import pandas as pd

import numpy as np

import pickle

m1="SepsisRF.sav"

m2="SevereRF.sav"

m3="Shock.sav"

Sep1=pickle.load(open(m1,'rb'))

Sep2=pickle.load(open(m2,'rb'))

Sep3=pickle.load(open(m3,'rb'))

Res=int(input("Enter the Respiratory Rate : "))

Hrt=int(input("Enter the Heart Rate : "))

Wbc=float(input("Enter the WBC count : "))

Temp=float(input("Enter the Temp : "))

cpr=float(input("enter the cpr"))

con=int(input("enter consious :1 or not:0"))

#Sepsis Prediction

s=Sep1.predict([[Res,Hrt,Wbc,Temp,cpr,con]])

if s[0]==1:

print("Sepsis Symptomes detected")

print("Let's Check the Stage")

Urine=float(input("Enter the Urine Output : "))

spo=int(input("Enter the SPO2 : "))

#Sep2=pickle.load(open(m2,'rb'))

d=Sep2.predict([[Urine,spo]])

#Severe Sepsis Prediction

if d[0]==2:

print("------------------------")

print("Severe Sepsis Symptomes detected")

#Septic Shock Prediction

print("Predicting chances for Septic shock ")

BPSys=int(input("Enter the BP-Systolic : "))

BPDia=int(input("Enter the Bp-Diastolic : "))

Urine = float(input("Enter the Urine Output : "))

Glucose=int(input("Enter the Glucose : "))

Creatinine=float(input("Enter the Creatinine : "))

Lactate=int(input("Enter the Lactate : "))

Bilirubin = float(input("Enter the Bilirubin rate : "))

Inr=float(input("Enter the Inr : "))

Platelets=int(input("Enter the Platelets Rate : "))

f=Sep3.predict([[BPSys,BPDia,Urine,Glucose,Creatinine,Lactate,Bilirubin,Inr,Platelets]])

if f[0]==3:

print("------------------------")

print("Septic Shock may occur ")

else:

print("------------------------")

print("Not having chances for septic shock ")

else:

print("------------------------")

print("Sepsis Only detected")

else:

print("------------------------")

print("It's Normal !!!")

plt.show()

**5.6 SEPSIS MODEL**

import pandas as pd

from sklearn.ensemble import RandomForestClassifier

import numpy as np

import pickle

from sklearn.model\_selection import train\_test\_split

data1=pd.read\_csv('data/Sepsis.csv')

data2=pd.read\_csv('data/SevereSepsis.csv')

data3=pd.read\_csv('data/SepticShock.csv')

#Sepsis

X1=data1.drop(columns=['Class'])

Y1=data1['Class']

#SevereSepsis

X2=data2.drop(columns=['Class'])

Y2=data2['Class']

#SepticShock

X3=data3.drop(columns=['Class'])

Y3=data3['Class']

#Preprocessing

d={'Normal':0,'Severe Sepsis':2,'Septic Shock':3,'Sepsis':1}

Y1=Y1.map(d)

Y2=Y2.map(d)

Y3=Y3.map(d)

#Train\_test\_split

X1\_train,X1\_test,Y1\_train,Y1\_test=train\_test\_split(X1,Y1,test\_size=.10)

X2\_train,X2\_test,Y2\_train,Y2\_test=train\_test\_split(X2,Y2,test\_size=.10)

X3\_train,X3\_test,Y3\_train,Y3\_test=train\_test\_split(X3,Y3,test\_size=.10)

#Classifier Sepsis\_RF

cls1=RandomForestClassifier()

cls1.fit(X1\_train,Y1\_train)

acc1=cls1.score(X1\_test,Y1\_test)

Filename = "SepsisRF.sav"

pickle.dump(cls1,open(Filename,'wb'))

#Classifier Sepsis\_Severe\_RF

cls2=RandomForestClassifier()

cls2.fit(X2\_train,Y2\_train)

acc2=cls2.score(X2\_test,Y2\_test)

Filename = "SevereRF.sav"

pickle.dump(cls2,open(Filename,'wb'))

#Classifier Septic\_shock

cls3=RandomForestClassifier()

cls3.fit(X3\_train,Y3\_train)

acc3=cls3.score(X3\_test,Y3\_test)

Filename = "Shock.sav"

pickle.dump(cls3,open(Filename,'wb'))

plt.show()

**6.7 FLASK CODE:**

from flask import Flask,render\_template,request,redirect

from sklearn.ensemble import RandomForestClassifier

import pickle

app=Flask(\_\_name\_\_)

@app.route('/')

def home():

return render\_template('blog-home.html')

@app.route('/predict\_sepsis',methods=['POST','GET'])

def sepsis():

if request.method=="POST":

model=pickle.load(open('test1.pickle','rb'))

d={"Yes":0,"No":1,"yes":0,"no":1,"YES":0,"NO":1}

Loc=request.form['Loc']

Res=int(request.form['Res'])

Hr=int(request.form['Hr'])

Wbc=int(request.form['Wbc'])

Temp=int(request.form['Temp'])

Con=request.form['Con']

Con=d[Con]

Crp=int(request.form['Crp'])

print(Crp,Loc,Res)

s=model.predict([[Res,Hr,Wbc,Temp,Con,Crp]])

if s[0]==0:

return render\_template('No\_Def.html')

else:

return redirect('/predict\_severe')

return render\_template('Predict\_sepsis.html')

@app.route('/predict\_severe',methods=['POST','GET'])

def severe():

if request.method=='POST':

model=pickle.load(open('test2.pickle','rb'))

un=float(request.form['Un'])

sp=int(request.form['SP'])

print(un,sp)

s=model.predict([[un,sp]])

print(s)

if s[0]==2:

return redirect('/predict\_shock')

else:

return render\_template('Only\_Sep.html')

return render\_template('Predict\_severe.html')

@app.route('/predict\_shock',methods=['POST','GET'])

def shock():

if request.method=='POST':

model=pickle.load(open('test3.pickle','rb'))

bp=int(request.form['BpS'])

bd=int(request.form['BpD'])

un=float(request.form['Un'])

gl=int(request.form['Gl'])

cr=int(request.form['Cr'])

lc=float(request.form['Lc'])

bl=float(request.form['Bl'])

inr=float(request.form['Inr'])

pl=float(request.form['PL'])

s=model.predict([[bp,bd,un,gl,cr,lc,bl,inr,pl]])

print(s)

if s[0]==3:

return "Septic Shock"

else:

return "Severe Sepsis"

return render\_template('Predict\_shock.html')

@app.route('/sepsis')

def sep():

return render\_template('Hospitals.html')

if \_\_name\_\_ == "\_\_main\_\_":

app.run(debug=True)

**6.8 UML DIAGRAMS**

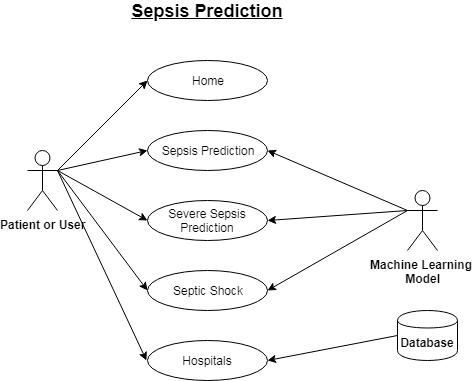
The Unified Modeling Language (UML) was created to forge a common, semantically and syntactically rich visual modeling language for the architecture, design, and implementation of complex software systems both structurally and behaviorally. UML has applications beyond software development, such as process flow in manufacturing. It is analogous to the blueprints used in other fields, and consists of different types of diagrams. In the aggregate, UML diagrams describe the boundary, structure, and the behavior of the system and the objects within it. UML is not a programming language but there are tools that can be used to generate code in various languages using UML diagrams. UML has a direct relation with object-oriented analysis and design.

There are many problem-solving paradigms or models in Computer Science, which relay on the study of algorithms and data. There are four problem-solving model categories: imperative, functional, declarative and object-oriented languages (OOP).  In object-oriented languages, algorithms are expressed by defining ‘objects’ and having the objects interact with each other. Those objects are things to be manipulated and they exist in the real world. They can be buildings, widgets on a desktop, or human beings.  Object-oriented languages dominate the programming world because they model real-world objects.

UML is a combination of several object-oriented notations: Object-Oriented Design, Object Modeling Technique, and Object-Oriented Software Engineering. UML uses the strengths of these three approaches to present a more consistent methodology that's easier to use. UML represents best practices for building and documenting different aspects of software and business system modeling

**6.8.1 USECASE DAIGRAM FOR SEPSIS PREDICTION**

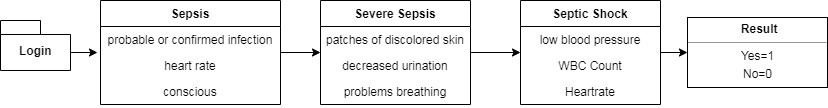
Use case diagrams are a set of use cases, actors, and their relationships. They represent the use case view of a system. A use case represents a particular functionality of a system. Hence, use case diagram is used to describe the relationships among the functionalities and their internal/external controllers. These controllers are known as actors.

****

**Fig 6.8.1**

**6.8.2 CLASS DIAGRAM FOR SEPSIS PREDICTION**

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram

****

**Fig 5.8.2**

**6.8.3 SEQUENCE DIAGRAM**

A sequence diagram is an interaction diagram. From the name, it is clear that the diagram deals with some sequences, which are the sequence of messages flowing from one object to another. Interaction among the components of a system is very important from implementation and execution perspective. Sequence diagram is used to visualize the sequence of calls in a system to perform a specific functionality.

**6.8.4ACTIVITY DIAGRAM**

Activity diagram describes the flow of control in a system. It consists of activities and links. The flow can be sequential, concurrent, or branched. Activities are nothing but the functions of a system. Numbers of activity diagrams are prepared to capture the entire flow in a system. Activity diagrams are used to visualize the flow of controls in a system. This is prepared to have an idea of how the system will work when executed.

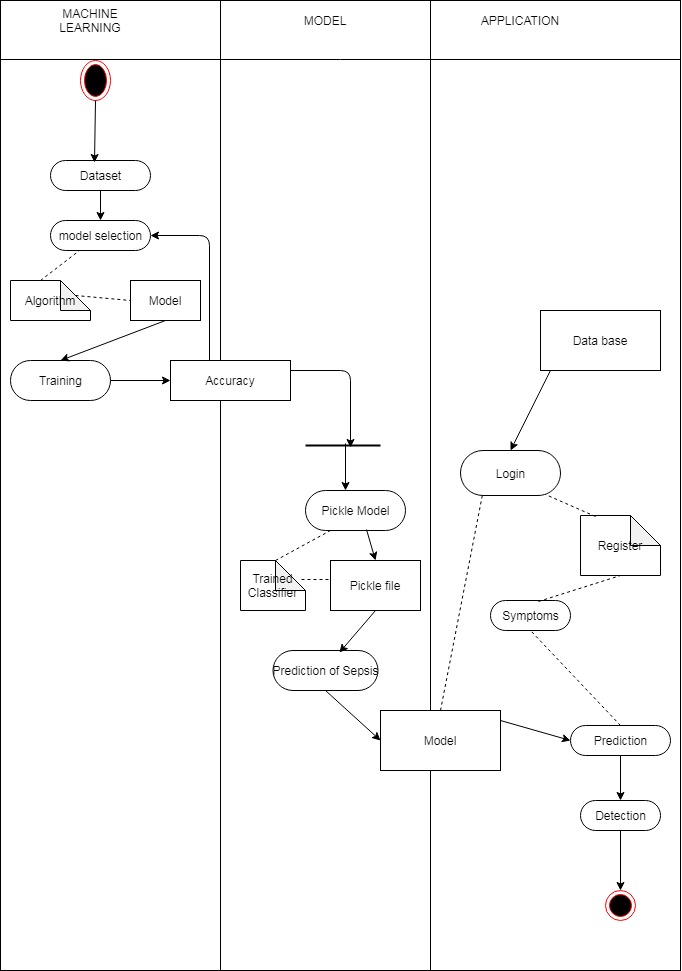
****

Fig 6.8.3

**6.8.5 DATAFLOW DIAGRAM**

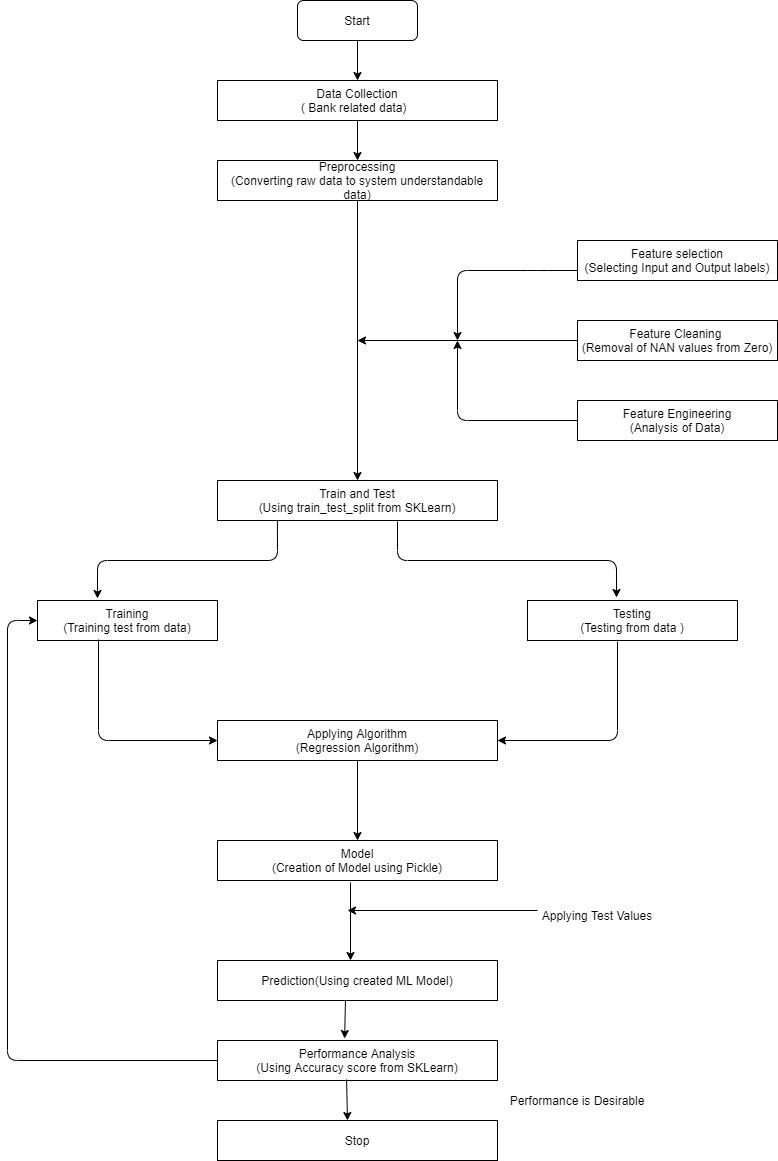


Fig 6.8.4

**CHAPTER 7**

**SYSTEM TESTING**

The aim of System Testing is to ensure that the System functions correctly and properly when all functions/features are bundled as a whole. Specific feature/function can be perfectly working on its own, but it might not when they are connected to each other. Imagine each function as a box and System is where all the boxes are linked and communicate to each other on the way they should. There is the risk that communication between boxes may not going well when you perform specific scenario and information need to flow among boxes.

**7.1 UNIT TESTING**

A unit test is the smallest testable part of an application like functions, classes, procedures, interfaces. Unit testing is a method by which individual units of source code are tested to determine if they are fit for use. Unit tests are basically written and executed by software developers to make sure that code meets its design and requirements and behaves as expected The goal of unit testing is to segregate each part of the program and test that the individual parts are working correctly .This means that for any function or procedure when a set of inputs are given then it should return the proper values. It should handle the failures gracefully during the course of execution when any invalid input is given .A unit test provides a written contract that the piece of code must assure. Hence it has several benefits. Unit testing is basically done before integration as shown in the image below.

**7.2 INTEGRATION TESTING**

Integration testing is a software testing methodology used to test individual software components or units of code to verify interaction between various software components and detect interface defects. Components are tested as a single group or organized in an iterative manner. After the integration testing has been performed on the components, they are readily available for system testing. Integration is a key software development life cycle SDLC strategy. Generally, small software systems are integrated and tested in a single phase, whereas larger systems involve several integration phases to build a complete system, such as integrating modules into low-level subsystems for integration with larger subsystems. Integration testing encompasses all aspects of a software system's performance, functionality and reliability. Most unit-tested software systems are comprised of integrated components that are tested for error isolation due to grouping. Module details are presumed accurate, but prior to integration testing, Each module is separately tested via partial component implementation.

**7.3 FUNCTIONAL TEST**

Functional testing is a type of testing which verifies that each function of the software application operates in conformance with the requirement specification. This testing mainly involves black box testing and it is not concerned about the source.

Valid : identified classes of valid input must be accepted

Invalid : identified classes of valid input must be rejected

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised

System/procedures : interfacing system or procedures must be invoked

**7.4 WHITE BOX TESTING**

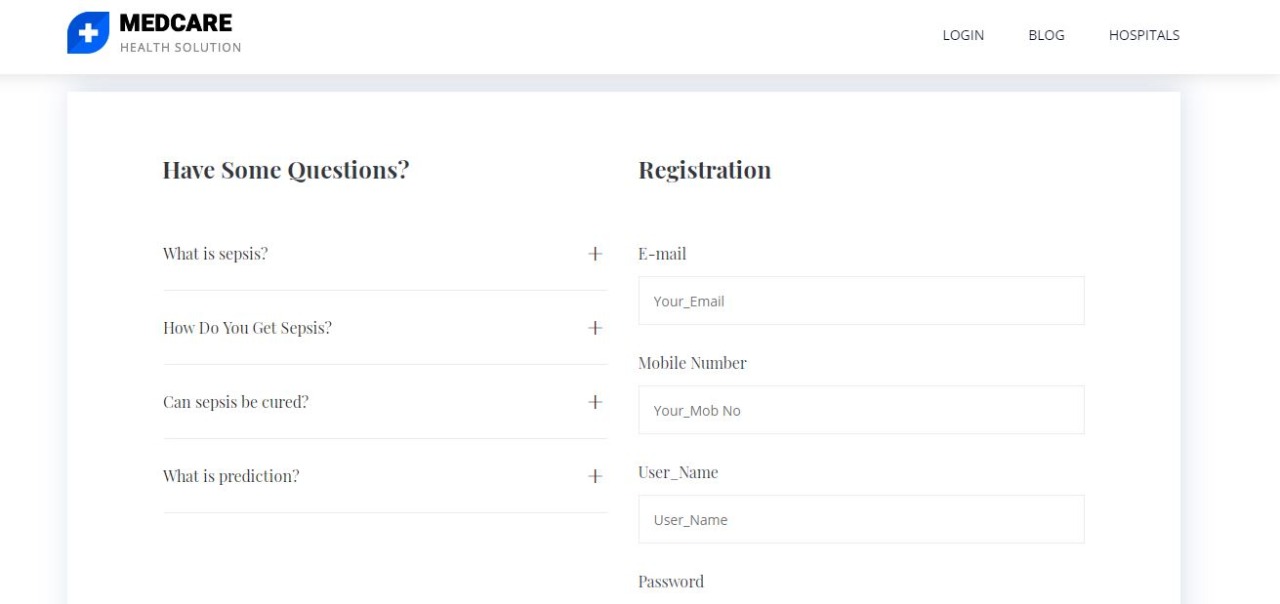
White box testing is a testing in which software tester has knowledge of the inner workings, structure and language of the software or atleast its purpose. It is used to test areas that can’t be reached from a black box level .

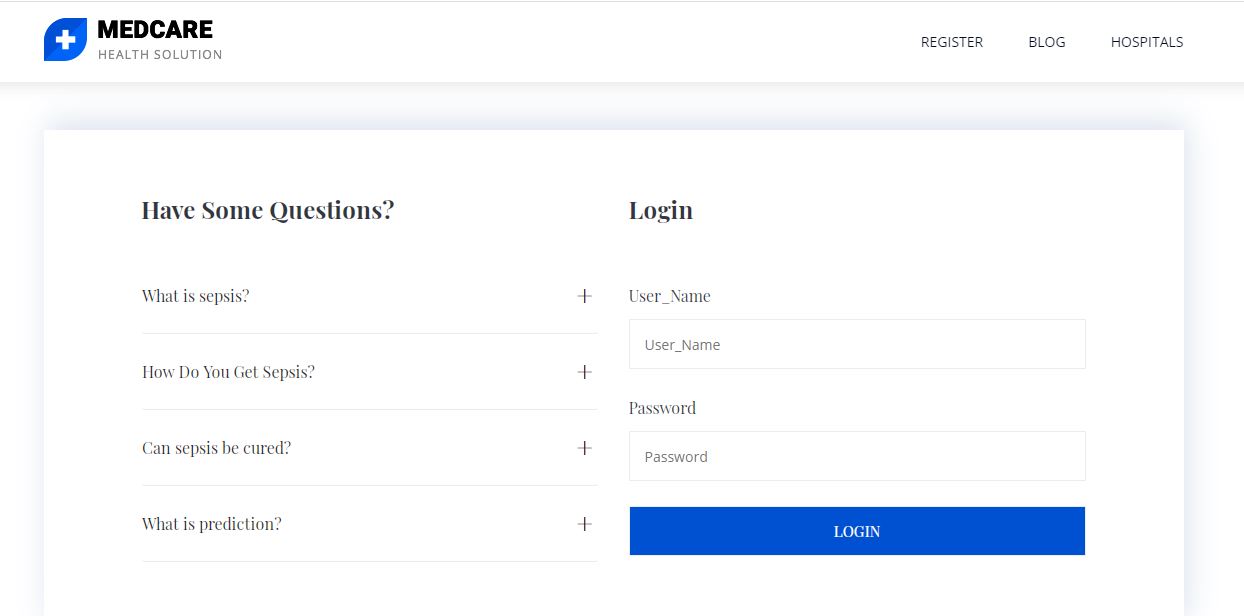
**7.5 BLACK BOX TESTING**

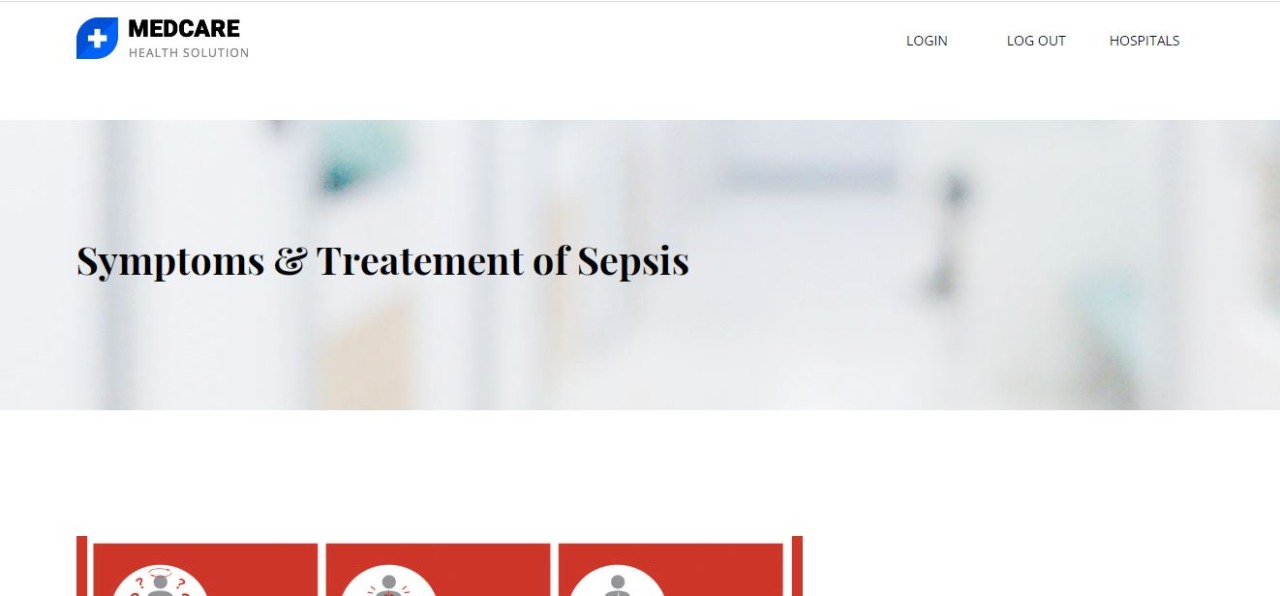
Black box testing is testing the software without any language of the inner workings . structure or language or the module being tested . black box tests , has most others kinds of test, must be return from a definitive source document , such as specification are requirements document, such as specification or requirements documentation.

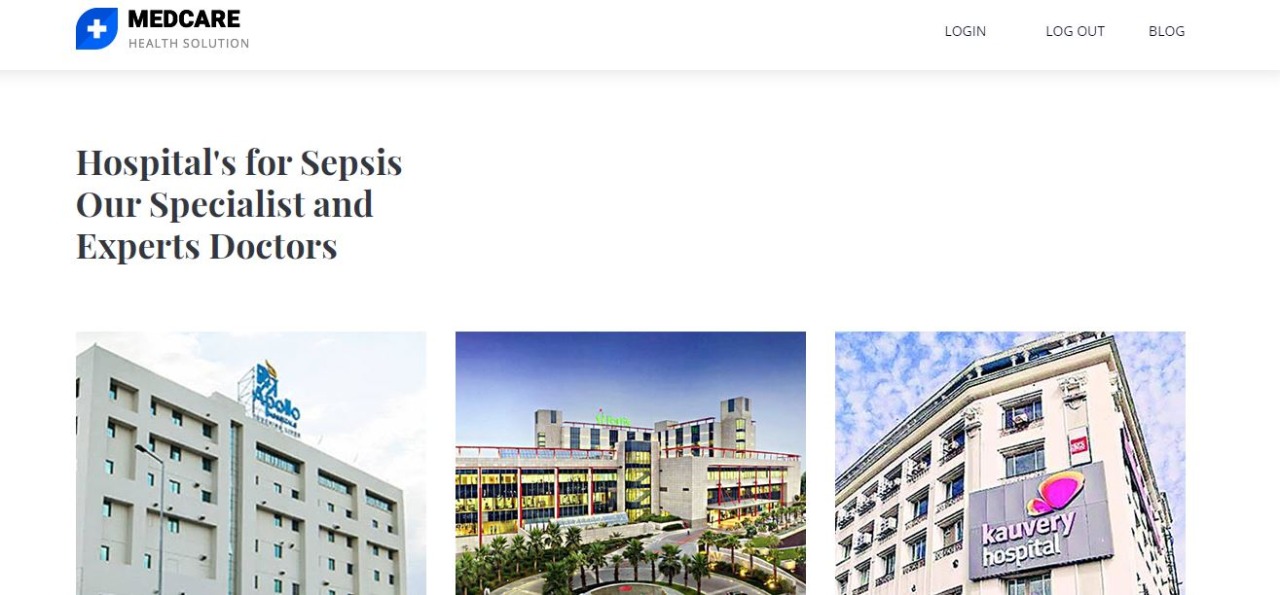
**SCREENSHOTS**

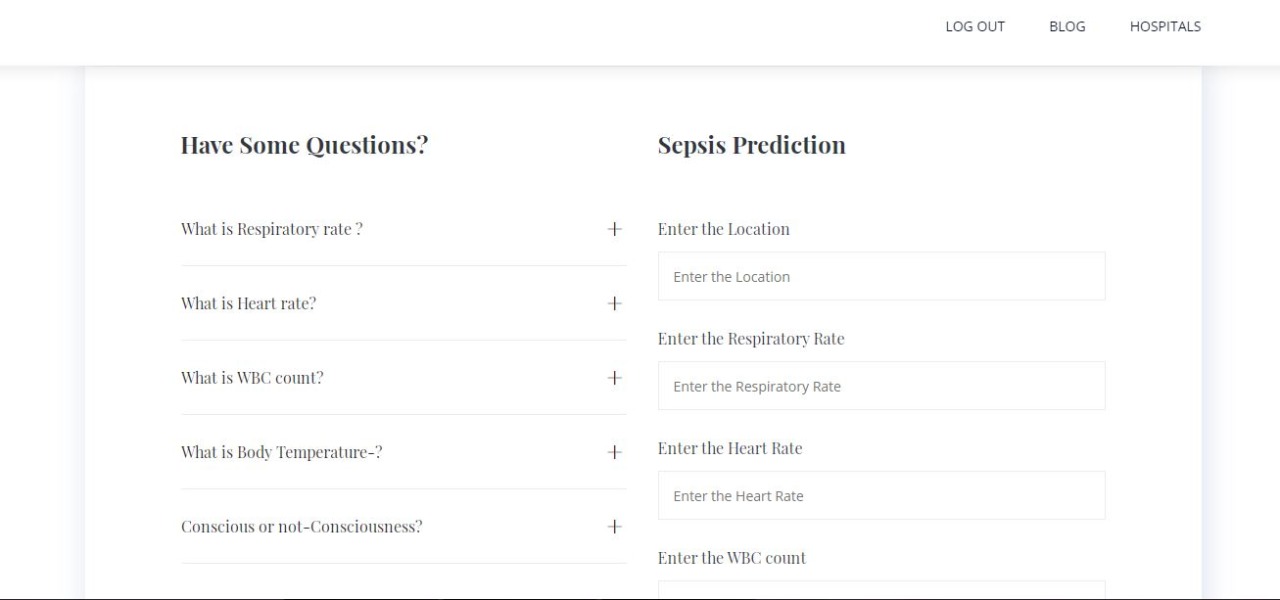
****

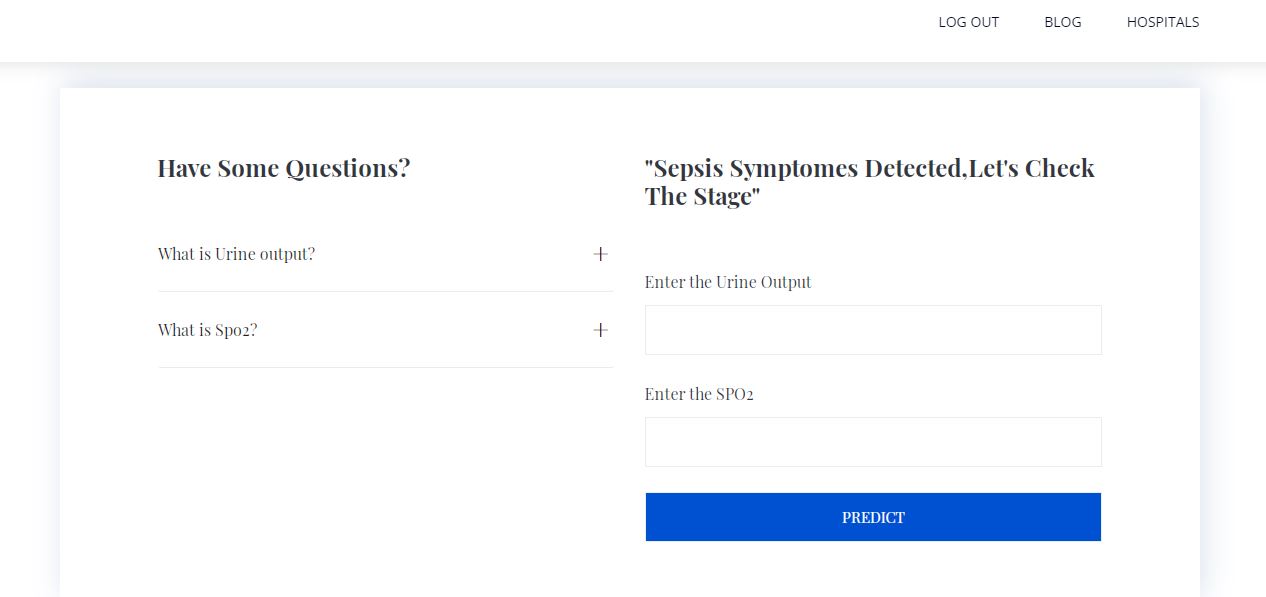
****

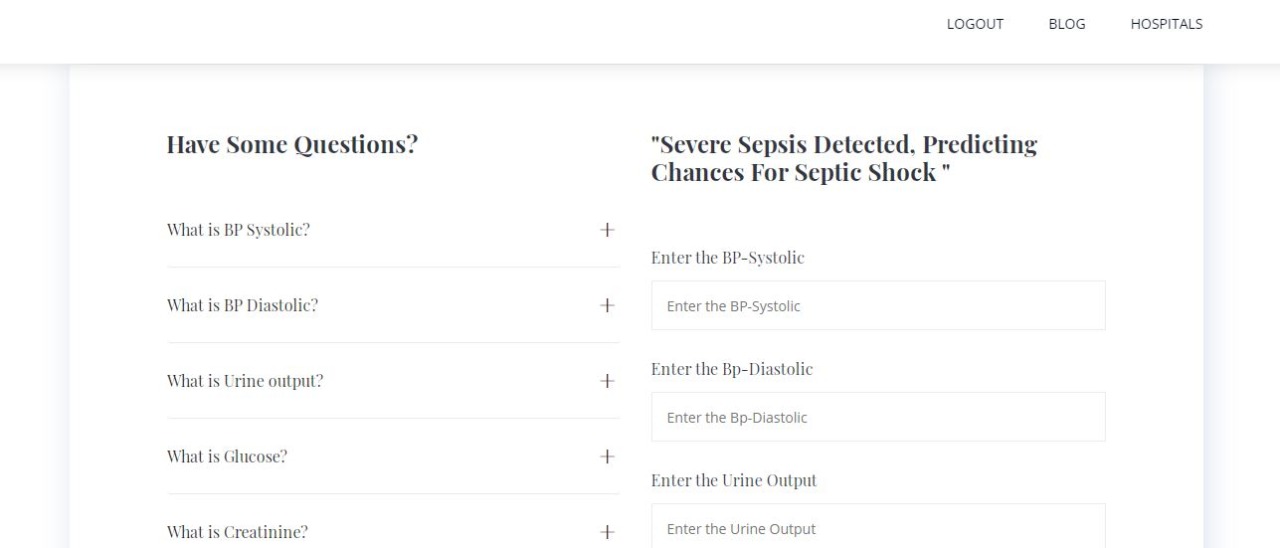
****

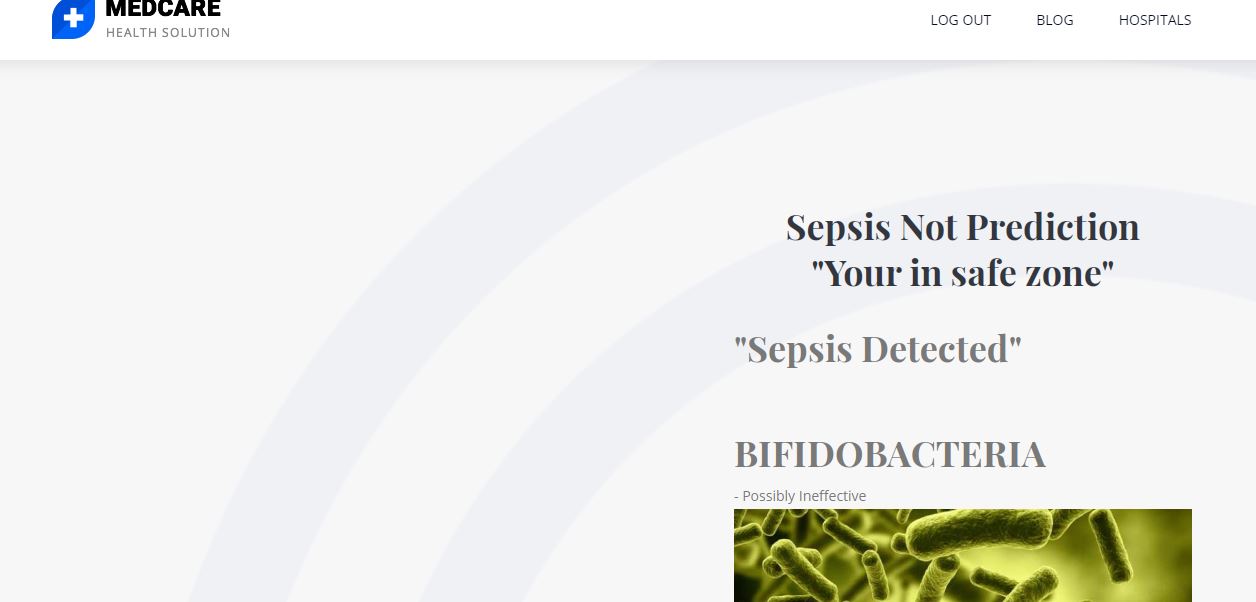
****

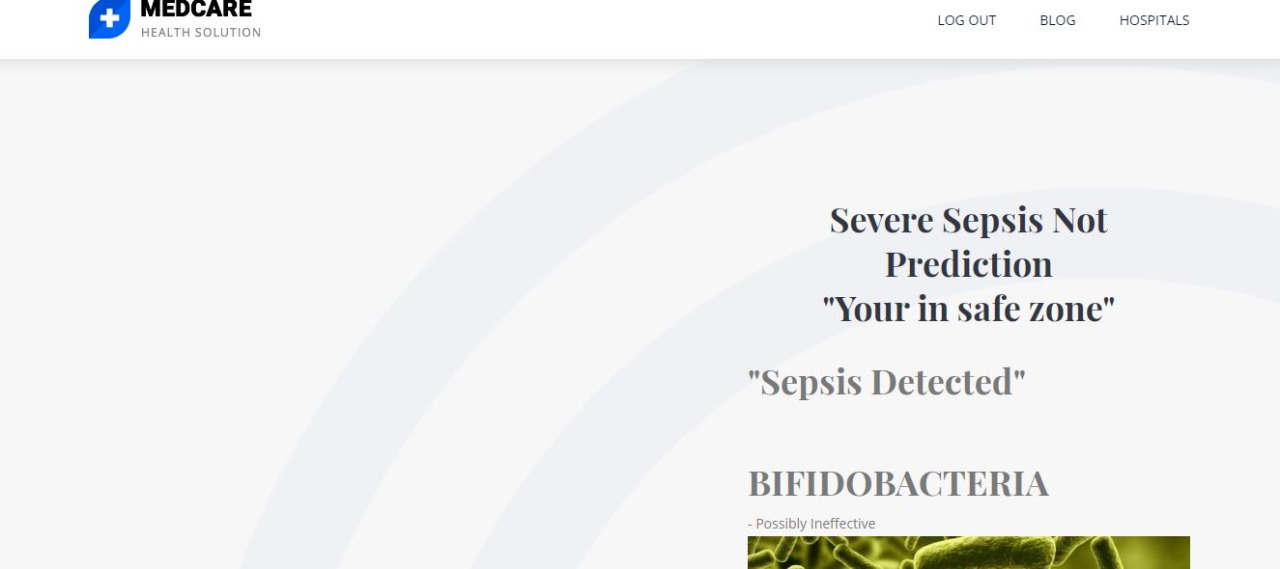
****

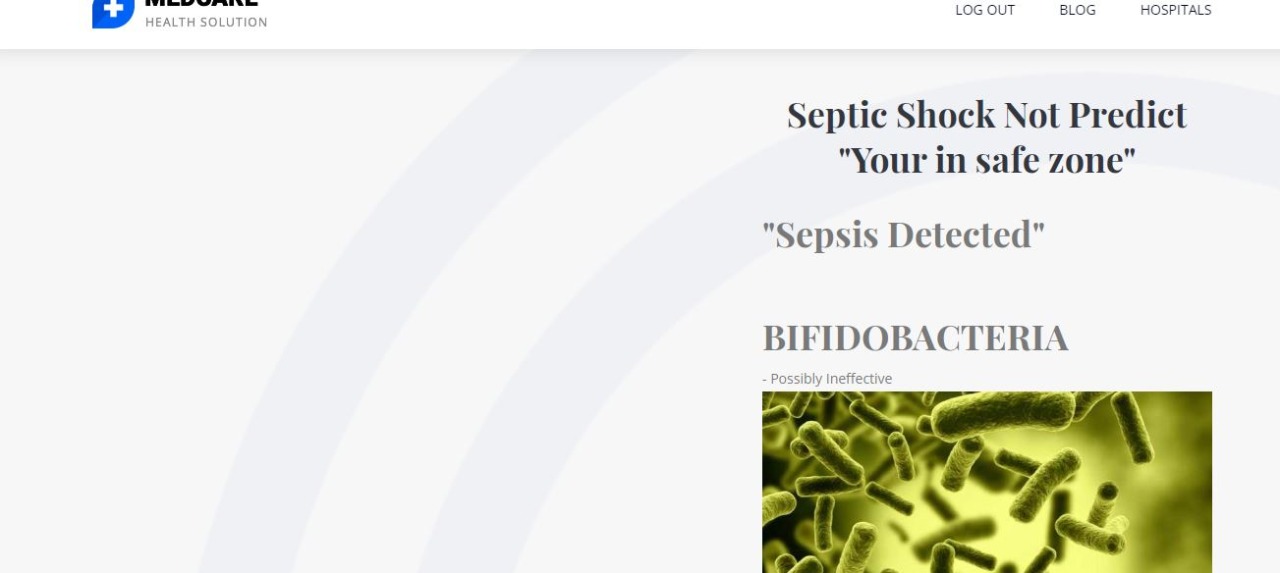
****

****

****

****

****

****

**CHAPTER 9**

**CONCLUSION**

Thus the proposed system ensures that the diagnosis of sepsis and the various phases of sepsis could be possibly achieved using random forest algorithm. Helps in early prediction of sepsis at different stages and could reduce death rate. In case the patient is identified in the third stage, applying the concept of Artificial Intelligence, the system provides suggestions of the hospitals where in the patient can be admitted immediately.

**REFERENCES**

[1] W. Raghupathi, and V Raghupathi. "Big data analytics in healthcare:promise and potential." Health information science and systems 2.1. 2014.

[2] D.Andaluz and R Ferrer. SIRS, qSOFA, and organ failure for assessingsepsis at the emergency department. Journal of thoracic disease, 9(6),1459. 2017.

[3] C. W. Seymour, V. X. Liu, T. J. Iwashyna, F. M. Brunkhorst, T. D. Rea,A. Scherag, and C. S. Deutschman. Assessment of clinical criteria forsepsis: for the Third International Consensus Definitions for Sepsis andSeptic Shock (Sepsis-3). Jama, 315(8), 762-774. 2016.

[4] M.Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D.Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.D. Chiche, C.M.Coopersmith, and R.S. Hotchkiss. The third international consensus

definitions for sepsis and septic shock (Sepsis-3). Jama, 315(8), pp.801-810. 2016.

[5] A Salcedo-Bernal, M. P. Villamil-Giraldo, and A. D. Moreno-Barbosa.Clinical Data Analysis: An opportunity to compare machine learningmethods. Procedia Computer Science, 100, 731-738. 2016.

[6] Y. Mao, W. Chen, Y. Chen, C. Lu, M. Kollef, and T. Bailey. An integrateddata mining approach to real-time clinical monitoring and deteriorationwarning. In Proceedings of the 18th ACM SIGKDD internationalconference on Knowledge discovery and data mining (pp. 1140-1148).ACM. 2012.

[7] F. Van Wyk, A. Khojandi, R.L. Davis, R. Kamaleswaran.Physiomarkers in Real-Time Continuous Physiological Data StreamsPredicts Adult Sepsis Onset Earlier Than Clinical Practice. 2018.doi:10.13140/RG.2.2.23671.52640/1

[8] Y Wang, and J. Patrick. Cascading classifiers for named entityrecognition in clinical notes. In Proceedings of the workshop onbiomedical information extraction (pp. 42-49). Association forComputational Linguistics. 2009.

[9] M. Cvach. Monitor alarm fatigue: an integrative review. Biomedicalinstrumentation & technology, 46(4), pp.268-277. 2012.

[10] C.M. Torio, and B.J Moore. National inpatient hospital costs: the mostexpensive conditions by payer, 2013: statistical brief# 204. 2016.

[11] K.E. Henry, D.N. Hager, P.J. Pronovost and S. Saria. A targeted real-timeearly warning score (TREWScore) for septic shock. Science translationalmedicine, 7(299), pp.299ra122-299ra122, 2015.

[12] E. Gultepe, J.P. Green, H. Nguyen, J. Adams, T. Albertson, and I.Tagkopoulos, 2013. From vital signs to clinical outcomes for patientswith sepsis: a machine learning basis for a clinical decision supportsystem. Journal of the American Medical Informatics Association, 21(2),pp.315-325.

[13] E.P. Raith, A.A. Udy, M. Bailey, S. McGloughlin, C. MacIsaac, R.Bellomo, D.V. Pilcher. Prognostic accuracy of the SOFA score, SIRScriteria, and qSOFA score for in-hospital mortality among adults withsuspected infection admitted to the intensive care unit. Jama, 317(3),pp.290-300. 2017.

[14] S. Nemati, A. Holder, F. Razmi, M.D. Stanley, G.D. Clifford, and T.G.Buchman. An Interpretable Machine Learning Model for AccuratePrediction of Sepsis in the ICU. Critical care medicine, 46(4), pp.547-553.2018

[15] S. P.Shashikumar, M.D. Stanley, I. Sadiq , Q. Li, A.Holder, G.D. Clifford,& S. Nemati, Early Sepsis Detection in Critical Care Patients UsingMultiscale Blood Pressure and Heart rate Dynamics. Journal ofElectrocardiology, 2017.

[16] F. van Wyk, A.Khojandi, R. Kamaleswaran, O. Akbilgic, S. Nemati, R.LDavis, How Much Data Should We Collect? A Case Study in SepsisDetection Using Deep Learning”. IEEE NIH Special Topics Conferenceon Healthcare Innovations and Point-of-Care Technologies. 2017.

[17] P. Viola, and M. Jones, Rapid object detection using a boosted cascade ofsimple features. In Computer Vision and Pattern Recognition, CVPR2001. Proceedings of the 2001 IEEE Computer Society Conference on(Vol. 1, pp. I-I). IEEE, 2001.