**VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY**

Department of Computer Engineering

Project Report on

Application of Machine Learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis

In partial fulfilment of the Fourth Year (Semester–VIII), Bachelor of Engineering

(B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2017-2018

Project Mentor
DR. GRESHA BHATIA
Deputy HOD

Submitted by
JUHI BHAGTANI (D17B, 07)
AISHWARYA CHANDAK (D17B, 11)
NEERAJ JETHNANI (D17B, 26)
MIHIR WAGLE (D17B, 73)

(2017-18)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that ***Mihir Wagle, Neeraj Jethnani, Juhi Bhagtani, Aishwarya Chandak*** of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily completed the project on "**Application of Machine Learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Dr. Gresha Bhatia*** in the year 2017-2018 .

This thesis/dissertation/project report entitled **Application of Machine Learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis** by ***Mihir Wagle, Neeraj Jethnani, Juhi Bhagtani, Aishwarya Chandak*** is approved for the degree of Bachelor of Engineering.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

APEX KIDNEY
FOUNDATION

11 October 2017

The HOD
Computer Engg Dept
VESIT
Mumbai

Dear Professor:

This is to state that 4 of your students, 1. Mihir Wagle, 2.Neeraj Jethnani 3. Aishwarya Chandak 4. Juhi Bhagtani, will be working on an academic research project in the healthcare domain, using the database of patients undergoing dialysis at Apex Kidney Foundation.

We are happy to have them conduct this research and will provide the necessary guidance and support.

Viswanath Billa

Dr.Viswanath Billa MD, DM
Nephrologist and Transplant Physician

Fargo House, Next to Spectra Motors, Ramchandra Lane Extn. Kanchpada, Malad (West). Mumbai - 400064
Tel : 022 - 28889118, 022 - 28809114. Mob : 9029069060
www.apexkidneyfoundation.org
E 25617 (M)

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled **Application of Machine Learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis** by ***Mihir Wagle, Neeraj Jethnani, Juhi Bhagtani, Aishwarya Chandak*** is approved for the degree of Bachelor of Engineering.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

--	--
(Signature)	(Signature)
--	--
Mihir Wagle -73	Neeraj Jethnani - 26
--	--
(Signature)	(Signature)
--	--
Juhi Bhagtani-07	Aishwarya Chandak-11

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Deputy HOD, Computer Engineering Department, V.E.S.I.T. **Dr. Gresha Bhatia** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Engineering Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be able to:-

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Abstract

Millions of patients worldwide suffer from Kidney failure and require dialysis. In most cases, dialysis is started after the kidney function of the patient falls below a threshold. In this scenario the patient’s kidney is essentially non functional. In order to conduct dialysis, native arteriovenous fistulas are constructed to increase blood flow in the superficial vein. Over time, as dialysis continues, the patient may suffer from hypertension and reduced vein function leading to the collapse of the fistula. The ultrasound doppler test for checking the state of the fistula are expensive and doing it again and again is not feasible. This report proposes a system which takes as input all the factors provided in real time by the dialysis machine and then uses the same to make a prediction on the state of a fistula, saving both time and money of the patient.

INDEX

Chapter 1
Introduction

In this chapter, we introduce our project by discussing our motivation behind choosing this topic for the project. We also mention our objectives in this project and the relevance of our work considering the absence of any such system. We finally discuss methodologies to be used.

Millions of patients worldwide suffer from Kidney failure and require dialysis. In most cases, dialysis is started after the kidney function of the patient falls below a threshold. In this scenario the patient’s kidney is essentially non functional[1]. In order to conduct dialysis, native arteriovenous fistulas are constructed to increase blood flow in the superficial vein, and hence facilitate dialysis. Over time, as dialysis continues, the patient may suffer from hypertension and reduced vein function leading to the collapse of the fistula. The ultrasound doppler test for checking the state of the fistula is expensive and doing it on a monthly basis can involve a large cost is not feasible for all dialysis patients. We explore the field of Chronic Kidney Diseases[2] and their effects on the human body to gain an insight into improving patient lives. We propose a system that uses the real time data being output by a dialysis machine, specifically the Nipro Surdial55 Plus, selects a few key parameters and generates attributes based on the same. These attributes are used alongside known classification by an ultrasound doppler test in supervised algorithms like random forest classifier, support vector machines and multilayer perceptrons to generate a model which can accurately predict the state of a fistula. This model is used to generate reports on a per session basis, which can be provided to both nephrologists and patients. This report allows doctors to recommend an ultrasound doppler test only if the health of the fistula is poor[3], saving both time and money for the patient. Also, since a bad fistula and hypertension[4] is spotted early on, a new one can be constructed without hampering the process of dialysis in any manner.

1.1 Motivation

Kidney failure, also known as renal failure or renal insufficiency, is a medical condition of impaired kidney function in which the kidneys fail to adequately filter metabolic wastes from the blood. The two main forms are acute kidney injury, which is often reversible with adequate treatment, and chronic kidney disease, which is often not reversible.

Chronic kidney disease (CKD)[2] is progressive loss in kidney function over a period of months or years. When a patient is diagnosed as having CKD, he needs to be put on dialysis, for which a native arteriovenous fistula needs to be constructed. Dialysis is generally done for four hours three times a week to ensure acceptable Quality of life for the patient. Over time, as dialysis is conducted, some patients might suffer from hypertension and reduced vein function which may eventually lead to the fistula failing. Developing another fistula takes time and during this period, the patient is inconvenienced by dialysis via a catheter. Constant ultrasound doppler scans[5] to check fistula health are not feasible.

1.2 Drawback of current system

Currently, the amount of data being output by the dialysis machines is view and go. I.e. It is not stored. It is not used except during dialysis. Therefore, in the current system, there is no prediction and most procedures follow an action response method where a new fistula is created once the old one fails catastrophically and hypertension is not treated proactively.

1.3 Problem Definition

The implementation of this project has two primary objectives.

The first objective is given the continuous data stream from a dialysis machine, we have to predict whether the patient is likely to suffer from hypertension due to dialysis in the short-term, the mid-term, the long-term or not at all.

The second objective is to use the continuous data stream from the dialysis machine to check the vein function and predict the lifetime of the fistula created for dialysis. This allows doctors to figure out the patient state without any expensive scans and if need be, create a new fistula before the current one fails so as to not inconvenience the patient with an catheter while the fistula develops.

1.4 Relevance of the project

This project will streamline procedures carried out by nephrologists, allowing them to proactively treat patients to stop them from entering a hypertensive state and keeping them informed as to the state of the fistula, to allow for the creation of a new one before the current one fails so as to not inconvenience a patient with a catheter as a stopgap measure. Doing the same is likely to improve the quality of life of the patient allowing them to lead a mostly normal life even while they undergo dialysis.

1.5 Methodology Used

Various parameters related to hypertension and dialysis will be identified and verified by multivariate analysis of patients with nephrologists. This will allow us to understand normalcy and also abnormalcy. Extensive datasets of these parameters will be obtained, and then machine learning algorithms will be applied for prediction. The predicted results will be re-verified and used for achieving optimum accuracy.

Chapter 2
Literature Survey

In the literature survey, we detail all the sources we have considered. This includes papers referred with their abstracts and the inferences we have drawn from each of them. We then describe what we learnt from books referred as well as interactions with subject experts. This survey allows us to list our observations, making things clearer.

2.1 Research Papers

2.1.1 Manoj Reddy and John Cho, 'Detecting Chronic Kidney Disease Using Machine Learning', Qatar Foundation Annual Research Conference Proceedings Vol. 2016, issue 1, March, 2016.

2.1.1.a - Summary: The goal of this paper was to use machine learning techniques and build a classification model that can predict if the patient is suffering from CKD. This was achieved with the help of various parameters such as age, blood pressure, specific gravity etc which were identified during the research. Through this work indications of possibilities if the patient is at the risk of kidney failure will be detected. Chronic kidney disease (CKD) refers to the loss of kidney functions over time which is primarily to filter blood. Based on its severity it can be classified into various stages with the later ones requiring regular dialysis or kidney transplant. Chronic kidney disease mostly affects patients suffering from the complications of diabetes or high blood pressure and hinders their ability to carry out day-to-day activities. In Qatar, due to the rapidly changing lifestyle there has been an increase in the number of patients suffering from CKD. According to Hamad Medical Corporation, about 13% of Qatar's population suffers from CKD, whereas the global prevalence is estimated to be around 8-16%. CKD can be detected at an early stage and can help at-risk patients from a complete kidney failure by simple tests that involve measuring blood pressure, serum creatinine and urine albumin.

2.1.1.b - Inference: Vague study suggesting trends in algorithm uses without much clear input parameter declaration. To fill the missing data, data processing was performed hence training set did not have 100% observed parameter values. 6 different classification algorithms were used to compare their results. They are: logistic regression, decision tree, SVM with a linear kernel, SVM with a RBF kernel, Random Forest Classifier and Adaboost [6].

Algorithm used	**Accuracy**
SVM with a linear kernel	98%
Random Forest Classifier	96%
Adaboost	95%
Logistic Regression	91%
Decision Tree	90%
SVM with a RBF kernel	60%

Table 2.1 Comparison of algorithms

2.1.2 - BV Ravindra, N Sriraam, M Geetha, 'Discovery of significant parameters in kidney dialysis data sets by K-means algorithm', IEEE, pages 452-454, Nov. 2014.

2.1.2.a - Summary: This paper used K-means clustering algorithm to use elicit knowledge about interaction between many of the important factors and survival of the patient undergoing dialysis. The important parameters for kidney dialysis such as creatinine, sodium, urea, Kt/V plays an important role in deciding the survival prediction of the patients as well as the need for undergoing kidney transplantation. Several attempts have been made to derive automated decision making procedure for earlier prediction. This preliminary study investigates the importance of clustering technique for identifying the influence of kidney dialysis parameters. The clustering procedure predicts the survival period of the patients who is undergoing the dialysis procedure.

2.2.1.b - Inference: K-means algorithm was used to determine the most important parameters for a person undergoing dialysis and how these parameters affect the survival duration of the patients. K- means algorithm was applied on datasets based on 3 criterias - Applying clustering to a part of dataset and treating the rest as training data, clustering based on age and gender. Different parameters were tested and classified in different priority ranges - low, normal and high. From all the three datasets it was observed that the priority range of creatinine was the most important factor and that the clusters with low and high mean creatinine levels had low survival rate and the ones with normal level have a better survival rate [7].

2.1.3 - Jeffrey Navarro Rojas, 'A Systematic Review on the Quality of Life among Hemodialysis Patients in the Middle East Countries', International Journal of Science and Research Vol. 6, issue 5, pages 532-542, May 2017.

2.1.3.a - Summary: This paper discusses about the quality of life as life saving and life altering in patients undergoing hemodialysis as a therapy for end-stage renal disease.

The paper states that degree of lifestyle change needs following prescribed diet/fluid limits and medications and managing symptom burdens depends considerably on the modality chosen, and affects patients' daily health-related quality of life and health-related quality of life is the impact of a chronic disease and its treatment on patients' perceptions of their own physical and mental function. It is found that among people on hemodialysis, QoL scores are both a critical outcome and a predictor of hospitalization and death. This paper focuses on QoL in middle eastern countries only.

2.1.3.b - Inference: Dialysis affects quality of life to a great extent. It's a renal replacement therapy and leads to changes in a person's physical and mental health. The QoL differs from person to person based on their social surroundings and cultural differences, and thus the doctors must use proper QoL tools for studies or treatments. Different observations from the different tools utilized,point that the tools measured QoL differently. More researches need to be done to culturally adapt QOL tools to make them relevant to other cultures, religions, and countries [8].

2.1.4 - Enes Celik, Muhammet Atalay, Adil Kondiloglu, 'The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning', International Journal of Intelligent Systems and Applications in Engineering Vol. 4, issue special issue-1, pages 27-31, Dec. 2016.

2.1.4.a - Summary:

This paper defines Chronic kidney disease and describes the way in which this is detected or diagnosed - an increase of urinary albumin excretion lasting more than three months or with significant reduction in a kidney functions. Chronic kidney disease can lead to complications such as high blood pressure, anemia, bone disease and cardiovascular disease. This paper aims at determining the factors that are decisive for early detection of chronic kidney disease, launching early patient's treatment processes, prevent complications resulting from the disease and predict of disease. It aims diagnosis and prediction of disease using the data set that composes of data of 250 patients with chronic kidney disease and 150 healthy people. First, the chronic kidney disease data is classified with machine learning algorithms and then training and test results are analysed. The estimation results of chronic kidney disease are then compared with similar data and studies.

2.1.4.b - Inference: Chronic kidney disease affects the functioning of kidney and it's early detection is important for health as well as monetary reasons. In a dataset of 250 CKD patients and 150 non-patients, Decision trees and Support Vector Machines were used for classification. Training and testing process if these algorithms were tested by creating 2 datasets. The size of the dataset plays an important role in classification. From the test done it was found that Decision Trees(97-100%) are more successful than SVM(96-97%) [9].

2.1.5 - Aaron Stern, Soumya Sachdeva, Rohit Kapoor, Jasjit Singh, Sarthak Sachdeva, 'High blood pressure in dialysis patients: cause, pathophysiology, influence on morbidity, mortality and management', Journal of clinical and diagnostic research: JCDR Vol. 8, issue 6,pages ME01, 20 Jun. 2014.

2.1.5.a - Summary: The paper states that based on recent guidelines the initiation of dialysis when symptoms and signs of kidney failure are present and not merely a decrease in GFR. The most common complication post dialysis is the occurrence of hypotension. However many dialysis patients are found to be hypertensive. This paper mentions the cause and pathophysiology of hypertension in dialysis patients and its management.

2.1.5.b - Inference: The most important way to manage lifestyles of dialysis patients is to attain dry weight. These patients have to undergo lifestyle changes and antihypertensives There is a lack of any cardiovascular events until blood pressure reaches 180 mm Hg. Patients with hypotension or low blood pressure have poor ventricular function which may be the cause of higher mortality, whereas higher risk hypertensive patients may not have survived to enter the study thus leading to a survival bias [10].

2.2 Book Review

2.2.1 Jeremy Levy, Edwina Brown, Anastasia Lawrence, 'Oxford Handbook of Dialysis', Oxford University Press, Feb. 2016.

A literature survey composed merely of papers detailing research is informative, but does not provide a complete picture of the dialysis process. This handbook explains the process of dialysis in detail alongside all its parameters with clear diagrams. It is clear and concise and helped increase our knowledge of dialysis quite a bit [11].

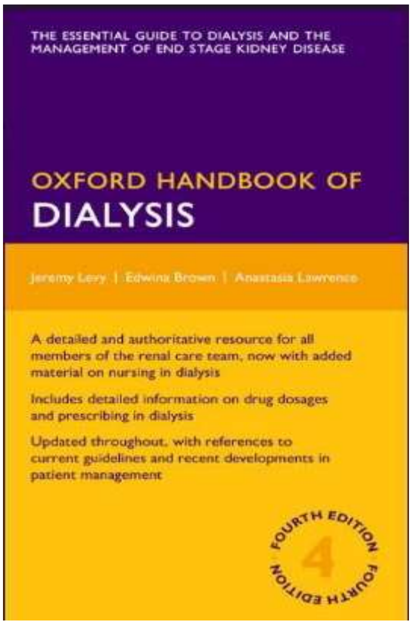


Fig 2.1

Oxford Handbook of Dialysis

http://oxfordmedicine.com/view/10.1093/med/9780199644766.001.0001/med-9780199644766

2.3- Interaction with subject expert

Multiple interactions with Dr. Viswanath Billa (MD, DM, nephrologist and kidney transplant physician, director at Apex Kidney Foundation)

Initially, we met Dr. Billa twice to explore any challenges in the field of nephrology which could benefit from the application of machine learning. These sessions served to help with defining our problem definition.

After defining the problem, we met Dr. Billa again to check whether our problem definition was feasible. This was followed by a lecture on the basics of dialysis and a suggestion to read the Oxford Handbook of Dialysis.

The next session was at the Apex Kidney Care dialysis centre in Sushrut Hospital, Chembur East. This time around, we got a look at the actual dialysis machines and how they work, including a query session with Dr. Deepa Usulumarty to answer our doubts.

In the next session we met Dr. Billa and Dr. Usulumarty (MD Internal medicine, fellowship in nephrology) to discuss the variables being output by the Nipro Surdial55 Plus Dialysis machine. We had another lecture to clear misconceptions and cleared the process of data collection.

Based on the data received we generated and analysed graphs on a per patient basis to identify significant parameters and verified the same with Dr. Billa and Dr. Usulumarty.

Chapter 3
Requirements of proposed system

In this chapter, we cover all the functional and non functional requirements, the constraints of the system and the hardware and tools to be used. We then have a project proposal, detailing activities in the project.

3.1 Functional Requirements

The functional requirements of a system are the technical details of a system which help define the functionality of the system and impose constraints on the design or implementation of the system.

1. Login for doctors/technicians - Having access control is essential to protect patient data which is of a sensitive nature.
2. Input section for adding patient variables. - Data will be input to the server as a csv file. This will be added to the database after training.
3. Clinical, biochemical and other factors which will be used for blood pressure prediction and vein function prediction modules.
4. Training section for feeding larger datasets - The training section will be modular with different models based on different algorithms.
5. Display section for output of models - Display section to show graphs of all models.
6. View of the patient details - The technicians/ doctors should be able to view the patient details and the generated prediction. This will be in the form of a generated report.
7. Graphical representation of results - Graphs of accuracy and other metrics must be imported into the system.
8. Delete patient details - In case a patient wishes to opt out of the system, his data must be removed
9. Results should be accurate - Accurate predictions are required for the system to be of use.
10. System should be efficient - The system must make efficient use of all available resources. RAM and GPU memory should be sufficient to hold dataset. Storage should be fast to reduce data transfer overhead.

3.2 Non Functional Requirements

Non functional requirement is a requirement that specifies the qualities of system on which the system shall be judged. Hence, it is often called as 'quality attributes', 'quality goals', 'quality of services requirements', 'constraints' and 'non-behavioural requirements'.

1. System should produce fast results. - Immediate outputs and transitions between different activities are expected from the system.
2. System should be portable. - Hybrid application will ensure portability for the system. It will be easily operated in multiple hardware and software specifications.
3. UI should be user friendly - UI will be designed by following common standards which are familiar for the masses. It will be concise and forgiving.
4. System should be extensible - Any further modification can be easily deployed to the system.
5. Security of the system. - Authentication will be provided to the system by password and also different privileges will be granted to different levels of users.

3.3 Constraints

1. The performance of the algorithm and hence the system will be limited by the sheer volume of data. Since the data involved is huge, significant time will go in pre-processing of data.
2. Depending on the pre-processing carried out, the algorithms to be used is limited.
3. Concentrates on a very specific ailment – fistula health. Does not take into account any other ailments
4. Current focus only on the major parameters as discovered through the graphical analysis.

3.4 Hardware Requirements

1. 2 Intel E5-2670v3 12 Core/24 Thread 2.3Ghz 30Mb Cache Processors
2. 128 GB RAM
3. Nvidia Quadro P600 - multiple units (minimum 4)
4. Internal drives - 1024GB Samsung 960 Pro NVMe SSD

5. External storage - Distributed Ceph Storage with SSD Caching Tier using 800GB Intel SSD DC S3610 Series Drives
6. Nipro Surdial55 Plus dialysis machine

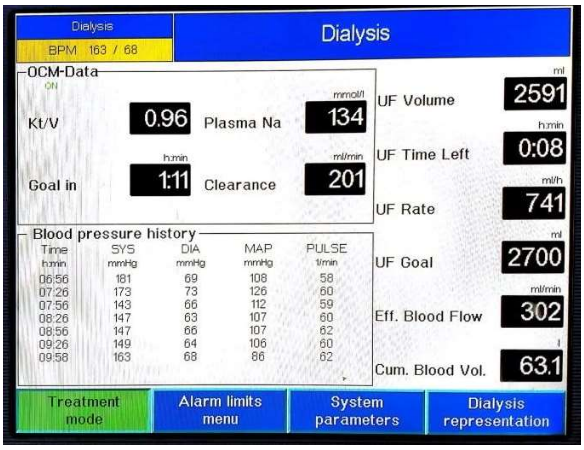


Fig 3.1: Treatment mode of Nipro Surdial55 Plus dialysis machine

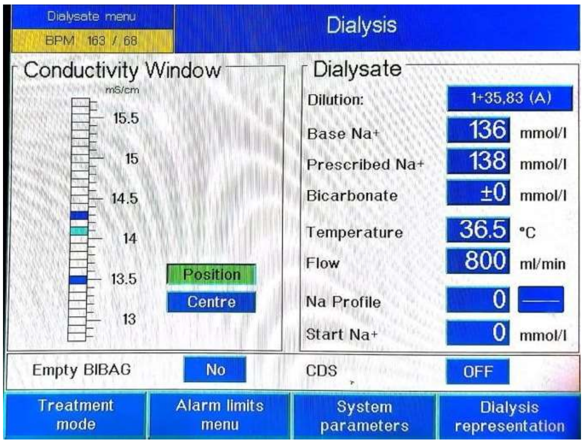


Fig 3.2: Dialysis representation of Nipro Surdial55 Plus dialysis machine

In the above figures we see two screen grabs from the Nipro Surdial55 Plus dialysis machine. The device has a default treatment mode which is shown in the Fig 3.1 and others for modifying the levels of alarms to ring, system parameters as well as dialysis representation (Fig 3.2) which is shown in the second image.

3.5 Tools utilized for the proposed system

- Scikit Learn - It is an open-source software library for machine learning across a range of tasks. It is a symbolic math library, and also used as a system for building and training neural networks to detect and decipher patterns and correlations, analogous to human learning and reasoning [12].It is used for both research and production at Google.
- Hybrid Application Frameworks [13] - Hybrid application Frameworks are preferred over traditional native OS applications since they reduce work. Frameworks like Electron allow us to make the application using web technologies such as HTML, CSS, Javascript and AJAX and then generate code for the same for Desktop OS like Windows and MacOS. The performance degradation is insignificant since the model will reside on the server itself.
- Django, Javascript, CSS, HTML5

Chapter 4
Proposed Design

In this chapter, we propose the design of the system using several diagrams. We have a system diagram and an overall block diagram to define modules in the project. We follow this with Module diagrams and DATA FLOW DIAGRAMs till level 2. This is to be followed by an ER diagram of the datastore and a Gantt chart showing the approximate timeframe.

4.1 System Design

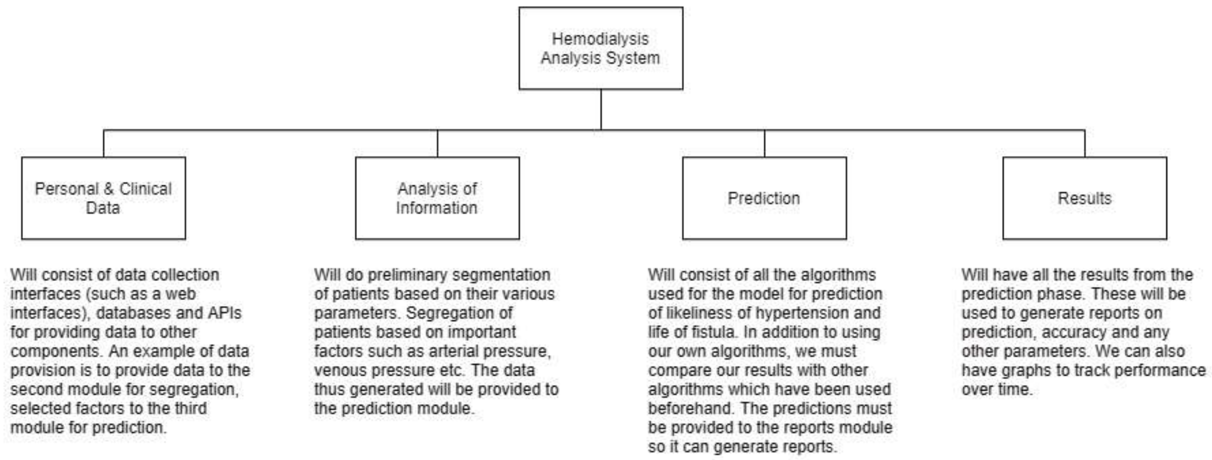


Fig 4.1- Architecture diagram

Architecture diagram shown in Fig 4.1 highlights different sub-systems to be built and their purpose. Different subsystems are Personal and Clinical Data interfaces, Analysis of Information, Prediction Model and Results. Personal and clinical data interface will be a data collection interface. Here technician and doctors will have a user login page. After login, user can upload data file of single dialysis session or multiple dialysis sessions into the system.

Analysis of collected data will be the second stage where preliminary segmentation of Clinical, Personal and Miscellaneous factors will be done based on the ranges.

Prediction will consist of all the algorithms used for the model for prediction of Likeliness of hypertension and Life of Fistula.

In the result phase, predicted results will be evaluated and compared with the known actual data of the patient and various reports like accuracy report, analysis report, efficiency report, prediction report, graphical representation of prediction will be generated.

4.2 Block Diagram

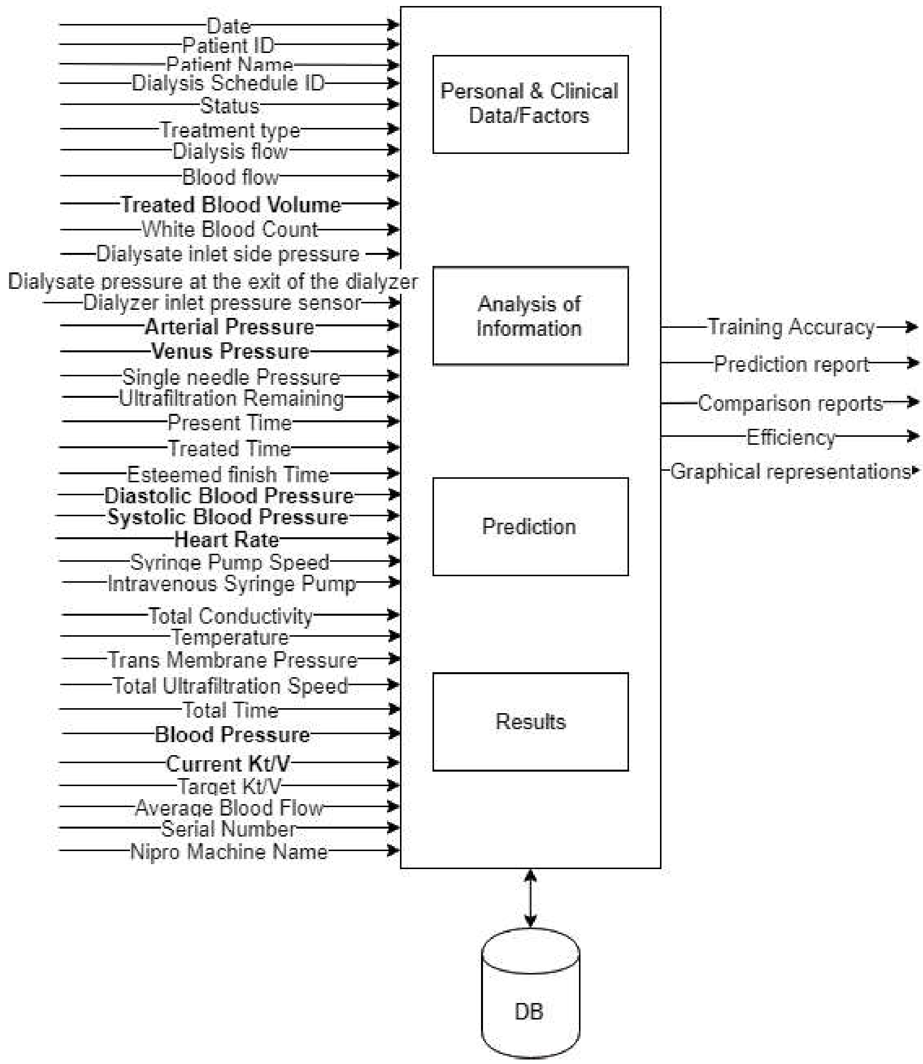


Fig 4.2- Block diagram

Block diagram shown in Fig 4.2 is an overview of the system along with input and expected output. The data input is to be taken from the Nipro Surdial55 Plus dialysis machine. While there are multiple parameters coming as input, only a few are directly involved in the system. They have been made bold to show significance.

4.3 Modular diagram

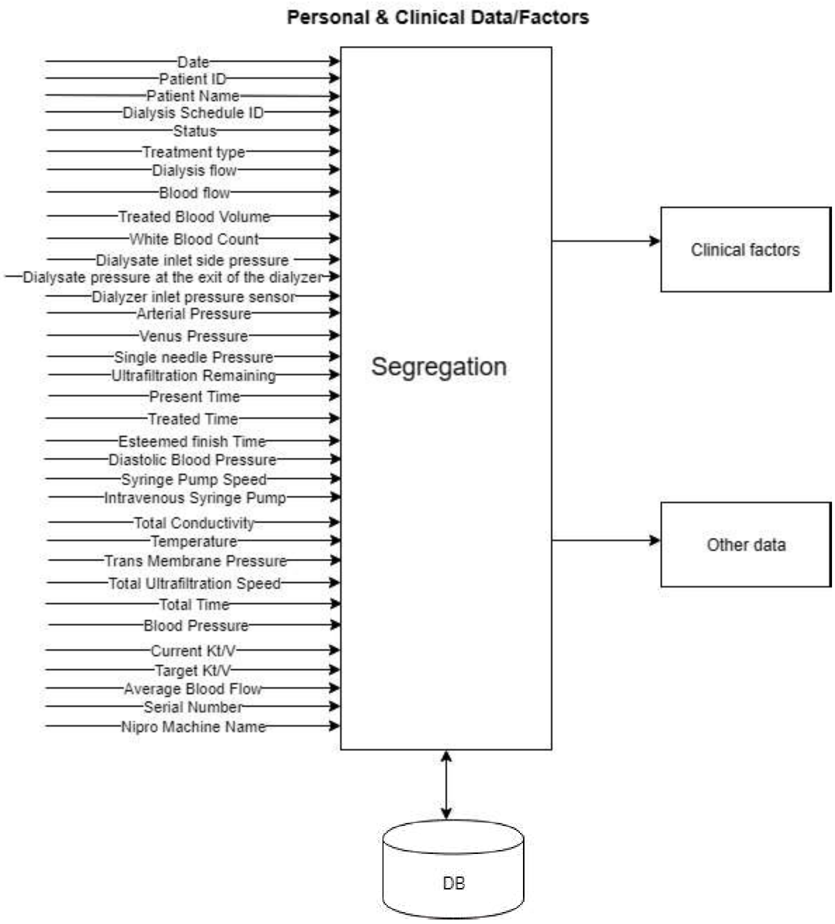


Fig 4.3-Personal & Clinical Data

Personal and clinical data module shown in Fig 4.3 consists of input factors from the dialysis machine and provides an output by categorizing these factors into clinical and other factors. All factors are not taken in to the final system but they are collected at this stage for analysis.

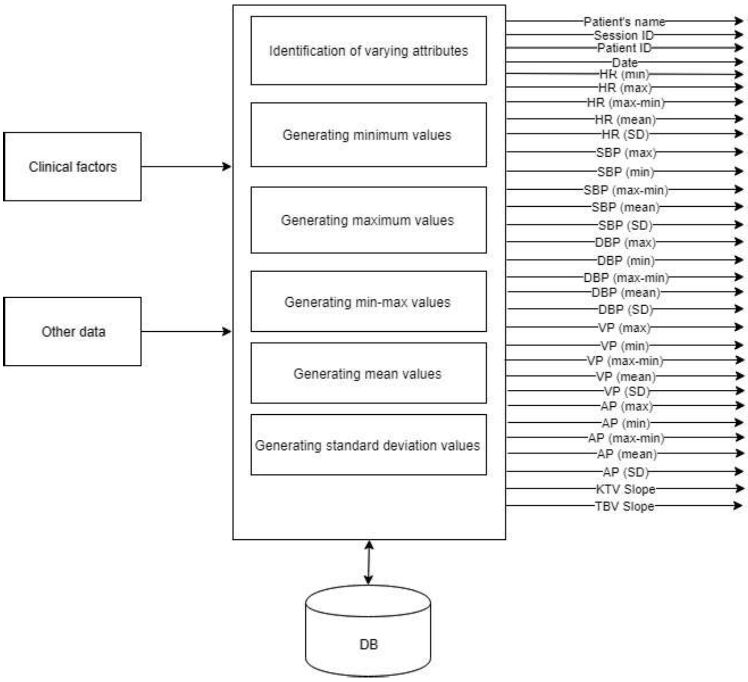


Fig 4.4 Analysis of Information

Output from the Segregation module is input for Analysis module as shown in Fig 4.4. In this module all factors are studied and a few important ones are identified. These factors are clustered into classes w.r.t. range i.e. max, min, standard deviation and mean. Possible combinations of these classes is output of this module.

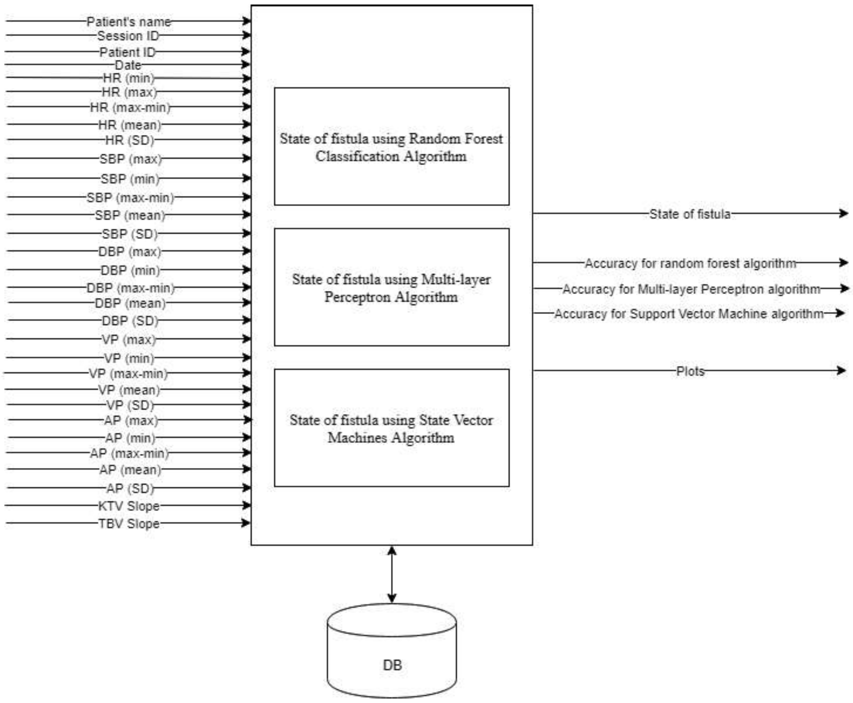


Fig 4.5 Prediction

Output from Analysis module is fed into Prediction Module shown in Fig 4.5. Here Machine Learning algorithms will be used to predict state of fistula, accuracy of algorithms and plots. The algorithms being used include Random Forest Classifiers, Support Vector Machines and Multi layer perceptrons.

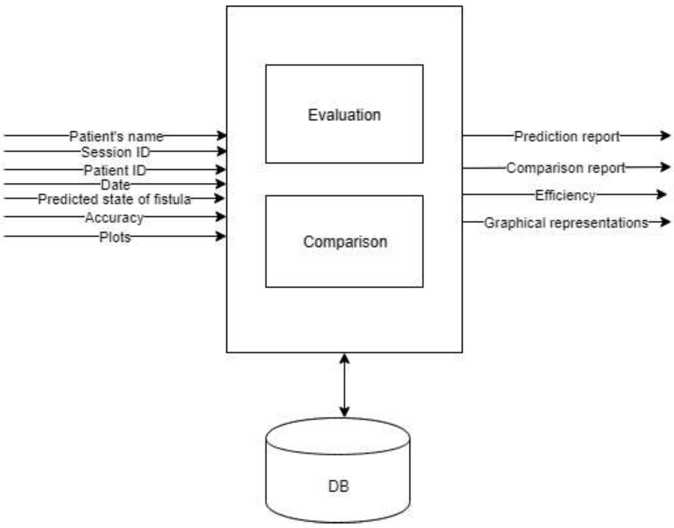


Fig 4.6 Results

In the Results module as in Fig 4.6, prediction will be compared with the known actual data of patient and report will be generated to evaluate the accuracy and efficiency of used algorithms along with the data of all the previous modules.

4.4 Design of system

4.4.a Data Flow Diagram

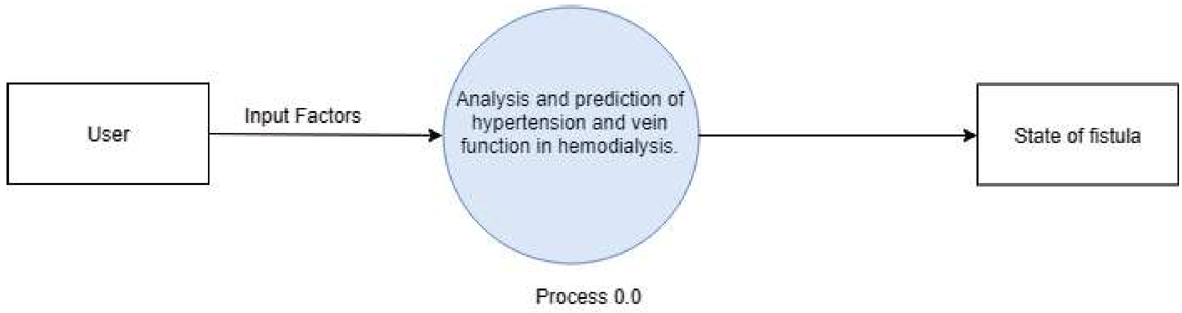


Fig 4.7 Level 0

Data Flow Diagram Level 0 shown in Fig 4.7 primary users will be doctors. The system is being built to predict if vein on which fistula is built will constrict with time, affecting dialysis performance and eventual failure of fistula.

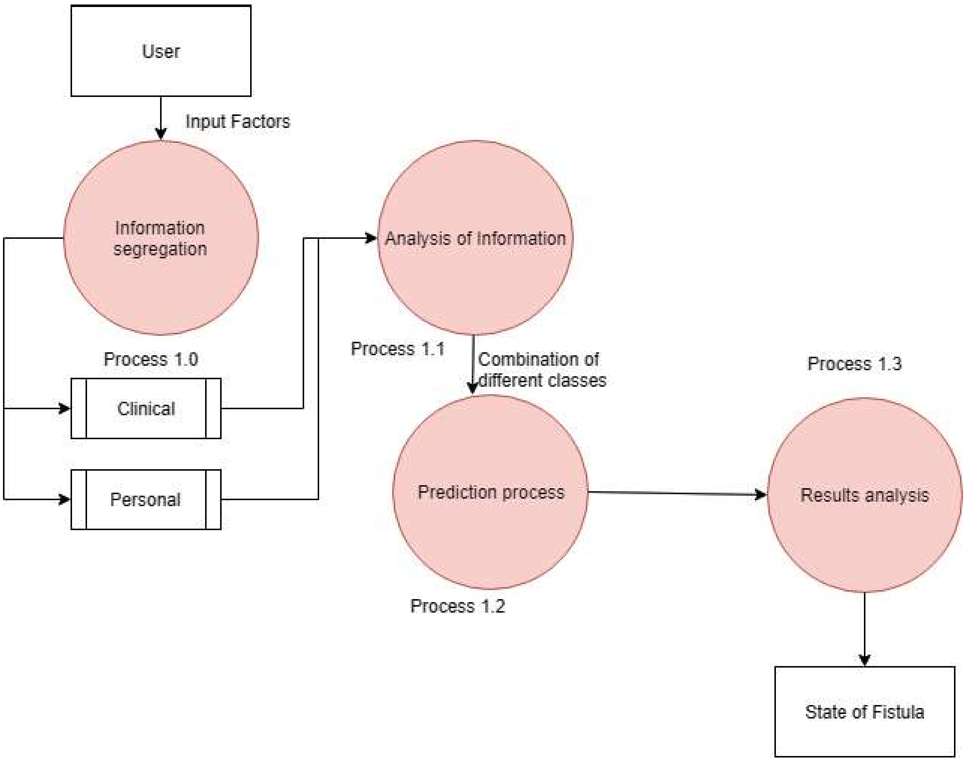


Fig 4.8 Level 1

Data Flow Diagram Level 1 shown in Fig 4.8 Input factors will be segmented into categories such as Clinical and Personal. After analysis of these factors, a few are selected and they are grouped into different classes and distinct possible combinations of these classes are fed into Prediction process.

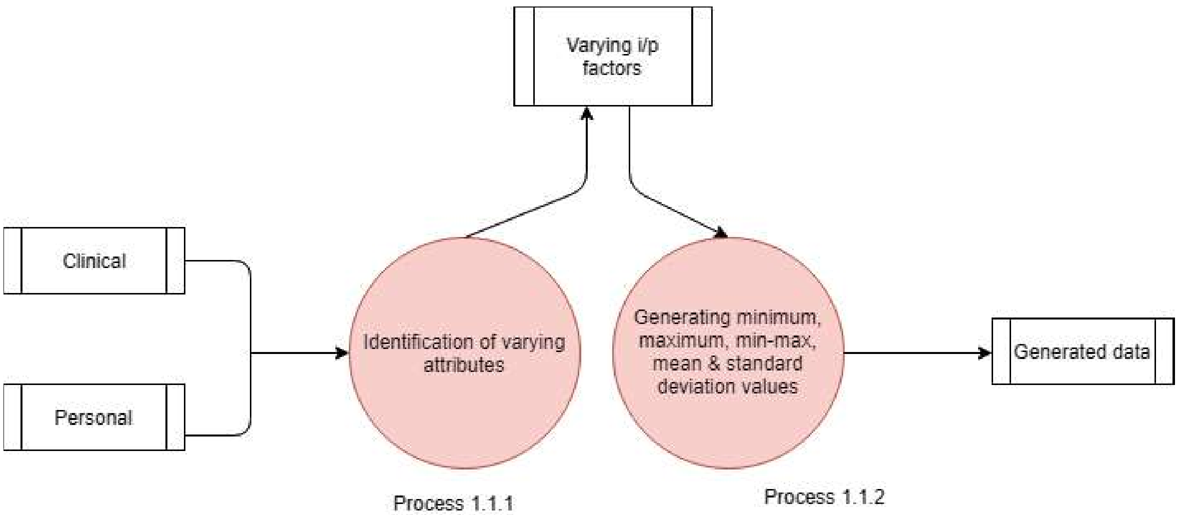


Fig 4.9 Level 2.1- Analysis

The Analysis module is expanded in Level 2.1 Data Flow Diagram shown in Fig 4.9. The segregated data is used to identify important parameters based on comparison via graphs. After this all the possible combinations of these factors are generated from this process.

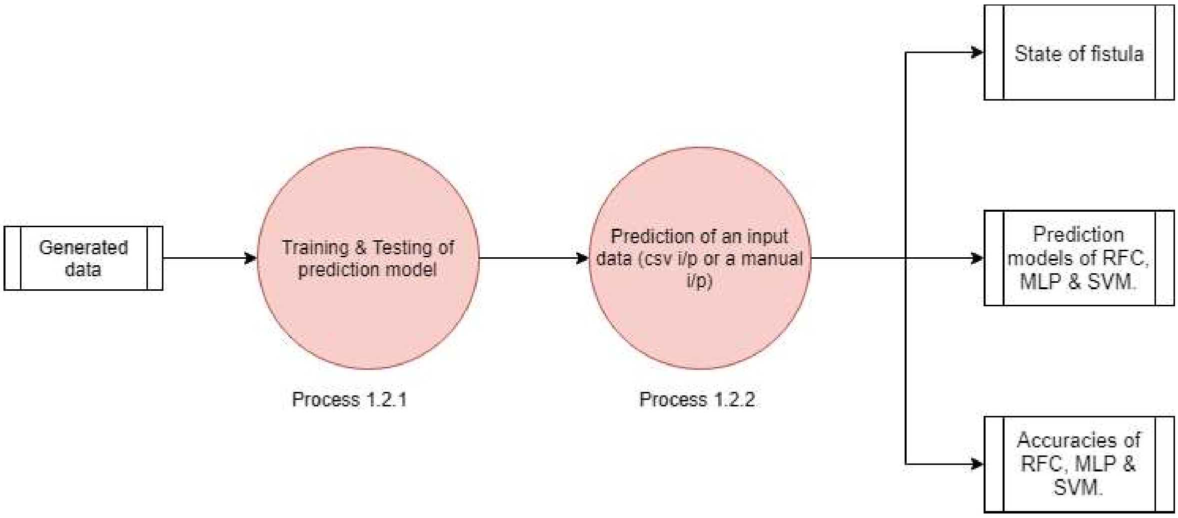


Fig 4.10 Level 2.2 - Prediction

Data Flow Diagram Level 2.2 shown in Fig 4.10 Prediction module is expanded in this level in which two processes are to train and test various machine learning models and generate prediction of the state of the fistula and accuracies of the algorithms used.

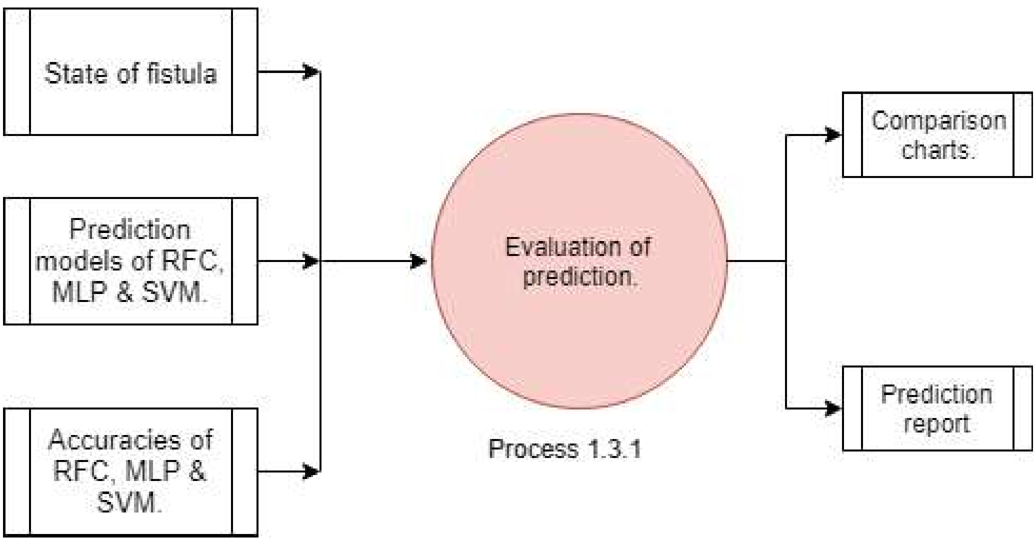


Fig 4.11 Level 2.3 - Results

Data Flow Diagram Level 2.3 shown in Fig 4.11 During results analysis the predicted outputs are evaluated with actual known conditions of patients. The process involved is Evaluation. Output of this process includes prediction reports and graphs for comparison.

4.4.b Flow of System

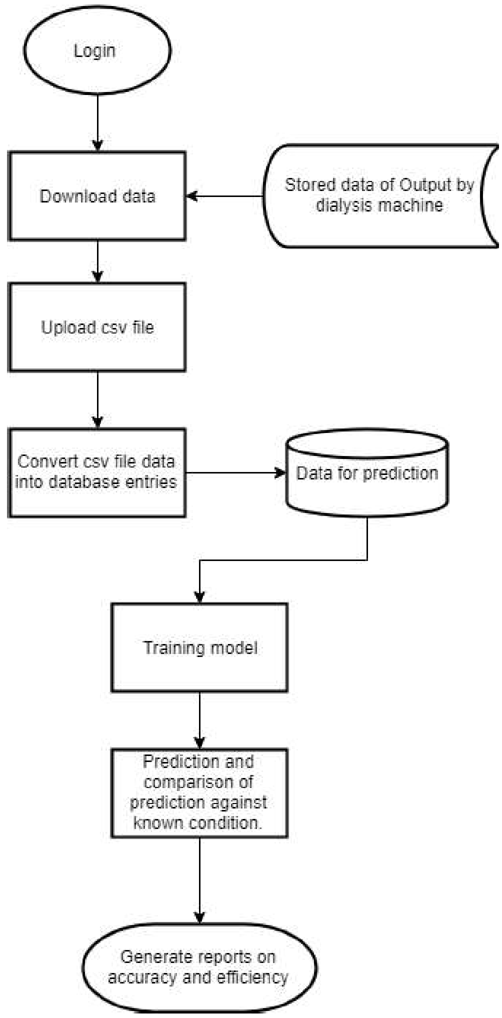


Fig. 4.12 Flow of System

The flow of the system is as follows in Fig 4.12:

a. Technician logs into the system.

b. Download of data output by the dialysis machine.

c. Upload of csv file to server.

d. Shell script to convert csv file to database entries.

e. Running script to add to database.

f. Add data to training set.

g. Use data to train model.

h. Test prediction against known patient condition. Optimise algorithms if necessary.

i. Generate reports.

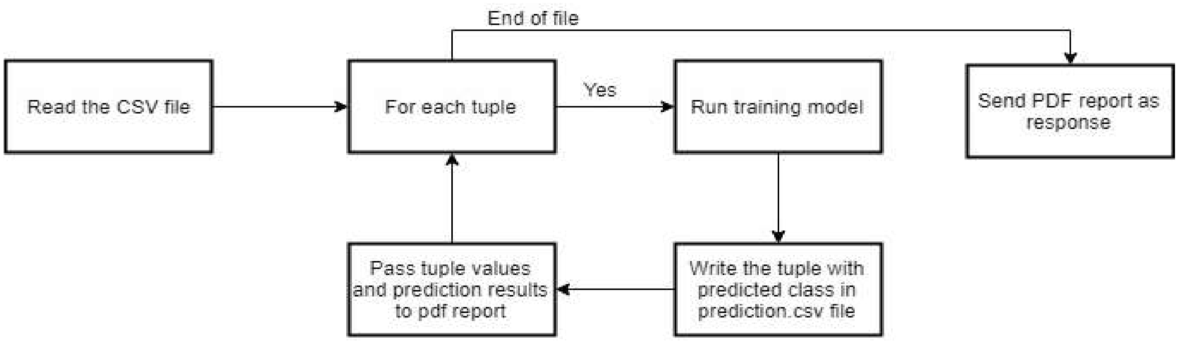


Fig 4.13 Flow for multiple patient prediction

Fig 4.13 shows the flow of the Prediction module. User uploads a CSV file. The CSV file must contain values for Session ID, patient ID, heart rate, systolic blood pressure, diastolic blood pressure, venous pressure, arterial pressure, total blood volume slope, KT/V slope for multiple patients. On submitting this file a downloadable report with prediction result for all the patients is generated.

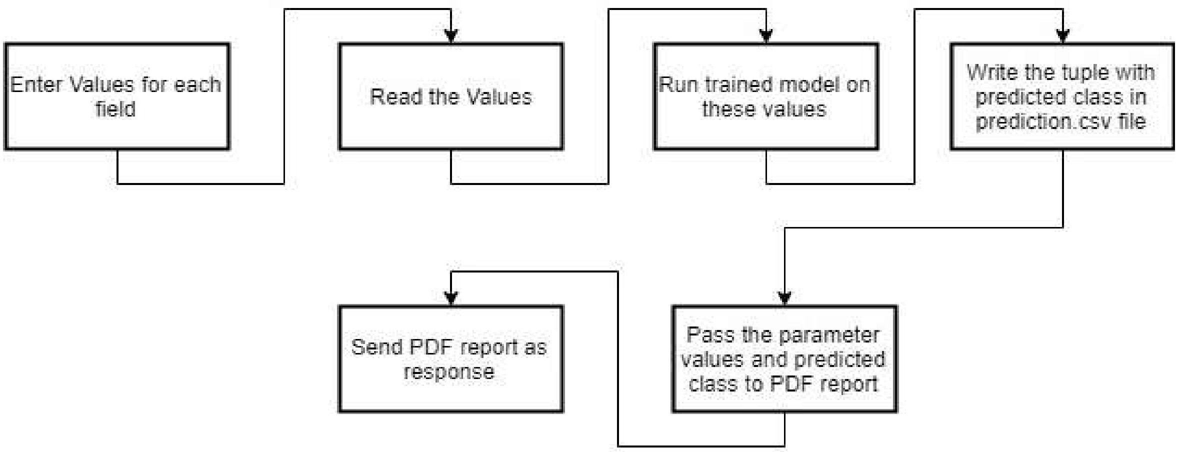


Fig 4.14 Flow for single patient prediction

The flowchart in Fig 4.14 depicts the flow of Single patient prediction module. In this, user has to manually enter values for each parameter required - Session ID, patient ID, heart rate, systolic blood pressure, diastolic blood pressure, venous pressure, arterial pressure, total blood volume slope, KT/V slope. On submitting these inputs a downloadable report with prediction result is generated.

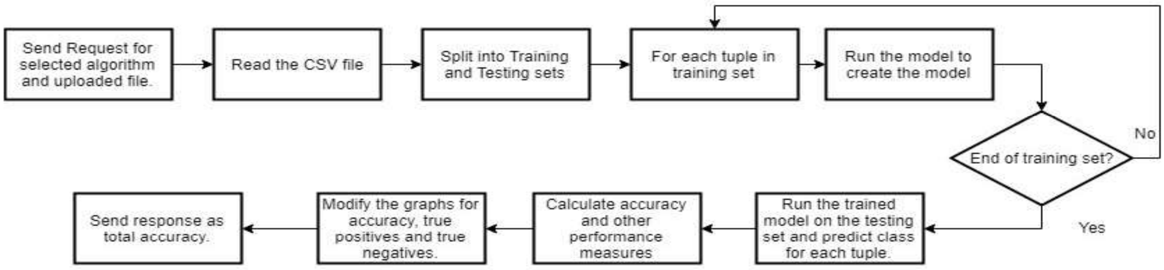


Fig 4.15 Flow for training and testing

The Fig 4.15 shows the flow for training and testing module. User chooses the Algorithm he wants to use for training. Then the user needs to upload the csv file for training and testing. The module splits the dataset provided into 80:20 ratio for training and testing respectively. Once the training is done, the generated model tests the values and produces an accuracy.

4.4.c ER diagram

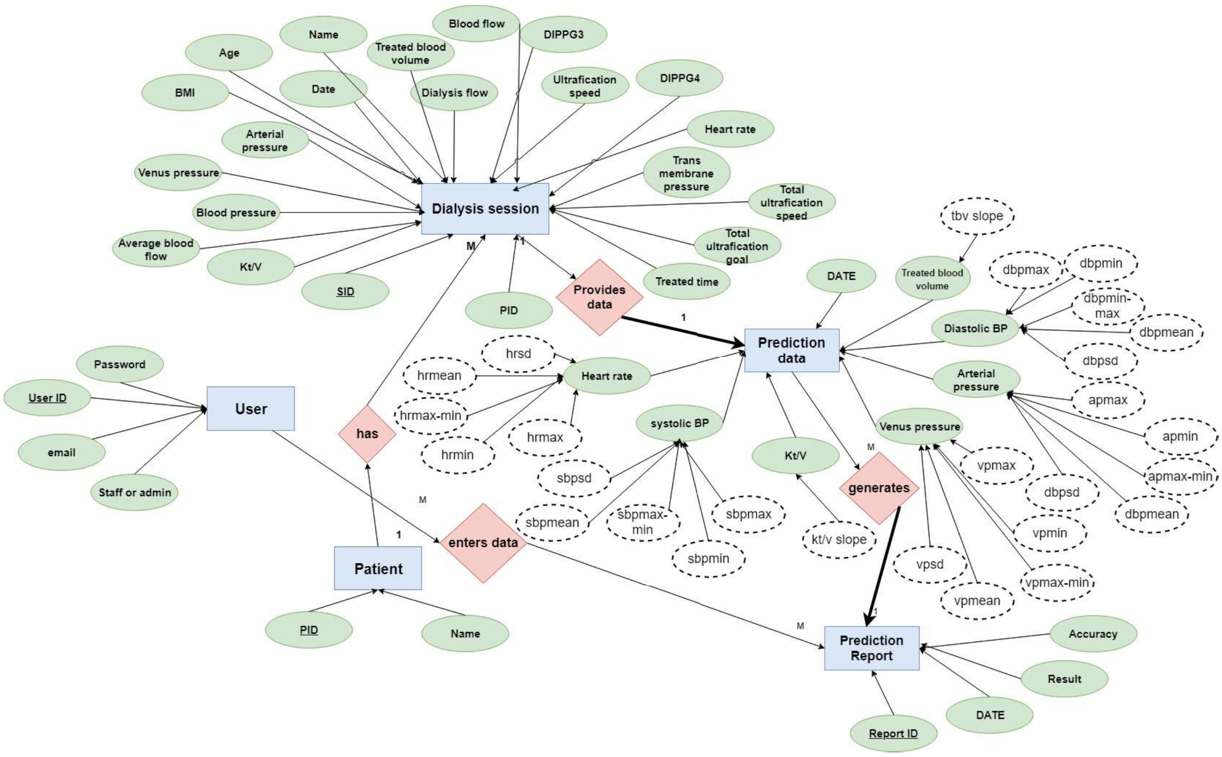


Fig 4.16 ER Diagram

In ER diagram in Fig 4.16:

1. Entities:
    a. Doctor: Authenticated user of the system.
    b. Patient: Patient's data is being used as input for the system.
    c. Dialysis session: Generates data for the system.
    d. Prediction: Output from Prediction algorithms are stored.
    e. Report: To provide measures for comparison.
2. Relationships:

. Patient has dialysis sessions which generates data.

a. Data generated from dialysis session is used for prediction.

b. Prediction model generates reports to keep check on accuracy and efficiency of the system.

4.5 Gantt Chart

Task Name	Status	Start Date	End Date	Durat...	Assigned To	% Complete
Project:Application of Machine Learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis.						
Project Definition	Complete	08/01/17	08/04/17	4d	MW,AC,NJ,JB	100%
Literature survey of paper	Complete	08/05/17	08/11/17	6d	MW,AC,NJ,JB	100%
Design of system	Complete	08/13/17	08/23/17	9d	MW,AC,NJ,JB	100%
Review 1	Complete	09/26/17	09/26/17	1d	MW,AC,NJ,JB	100%
Interaction with subject experts- Lectures on dialysis.	Complete	10/08/17	10/08/17	1d	MW,AC,NJ,JB	100%
Survey of relevant chapters in the Oxford handbook of dialysis.	Complete	10/08/17	10/11/17	4d	MW,AC,NJ,JB	100%
Interaction with subject experts- Demo of the dialysis process.	Complete	10/11/17	10/11/17	1d	MW,AC.NJ,JB	100%
Interaction with subject experts- discussion on next plan of action.	Complete	10/19/17	10/19/17	1d	MW,NJ,JB	100%
Paper 1	Complete	10/19/17	10/24/17	4d	MW,AC,NJ,JB	100%
Report submission	Complete	10/05/17	10/20/17	12d	MW,AC,NJ,JB	100%
Review 2	On Track	10/27/17	10/27/17	1d	MW,AC,NJ,JB	100%
Obtain 15 minute data from Apex Kidney Foundation.	Not Started	11/05/17	11/05/17	1d	MW,AC,NJ,JB	0%
Use it to define normalcy and define outliers.	Not Started	11/06/17	11/22/17	13d	MW,AC,NJ,JB	0%
Use it to define normalcy and define outliers.	Not Started	11/06/17	11/22/17	13d	MW,AC,NJ,JB	0%
Multivariate analysis to ascertain important factors.	Not Started	11/23/17	12/08/17	12d	MW,AC,NJ,JB	0%
Use design algorithm to predict the life of the fistula.	Not Started	12/11/17	12/19/17	7d	MW,AC,NJ,JB	0%
Obtain 1 second data from Apex Kidney Foundation.	Not Started	12/12/17	12/12/17	1d	MW,AC,NJ,JB	0%
Use data to train a model.	Not Started	12/15/17	01/12/18	21d	MW,AC,NJ,JB	0%
Compare results with an ultrasound doppler test.	Not Started	01/16/18	01/19/18	4d	MW,AC,NJ,JB	0%
Modify model to increase accuracy if needed.	Not Started	01/20/18	02/07/18	14d	MW,AC,NJ,JB	0%

Fig 4.17

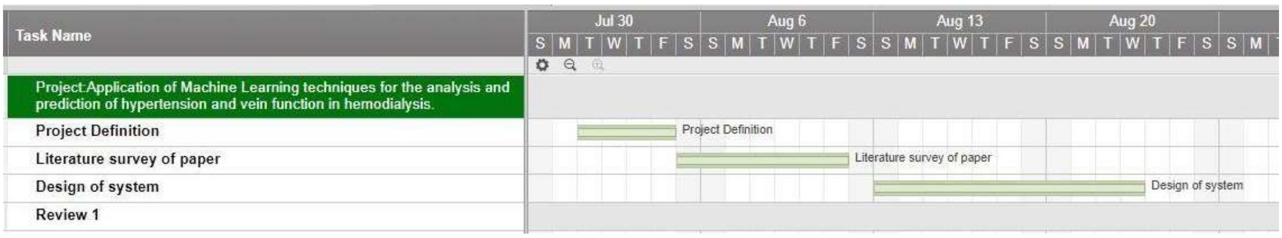


Fig 4.17 Part 1

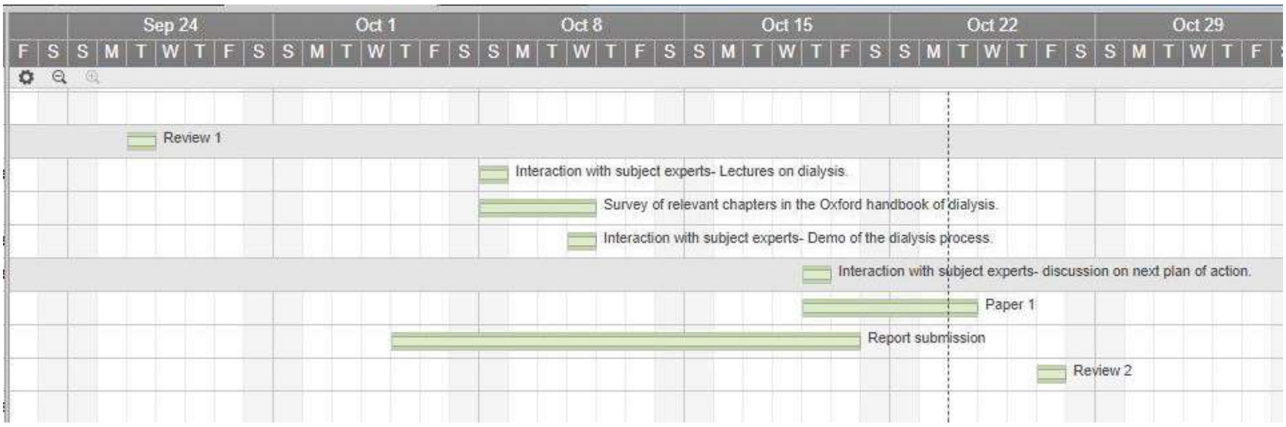


Fig 4.17 Part 2

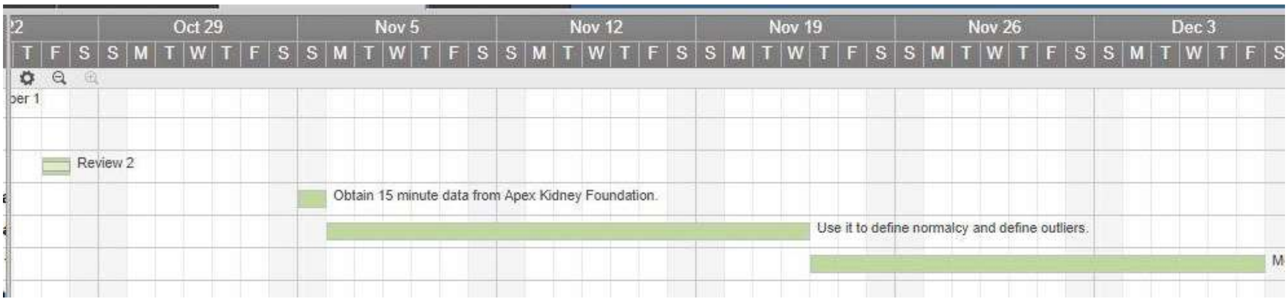


Fig 4.17 Part 3

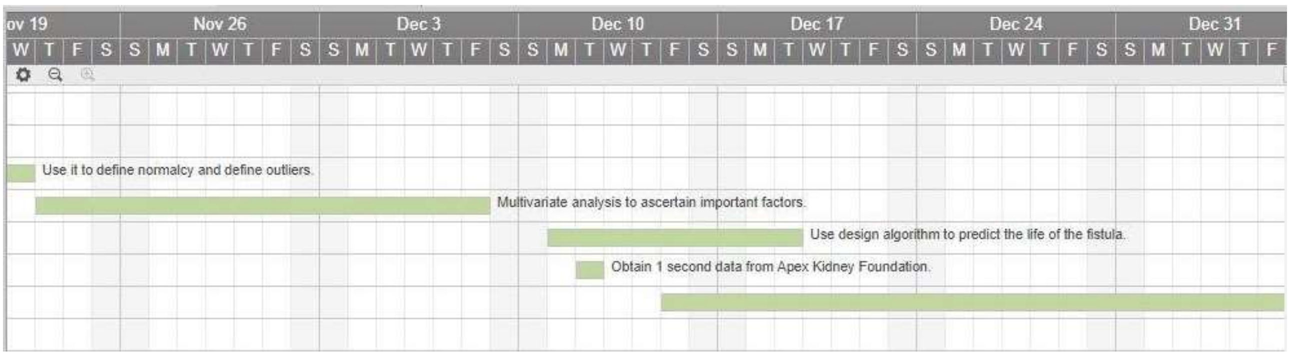


Fig 4.17 Part 4

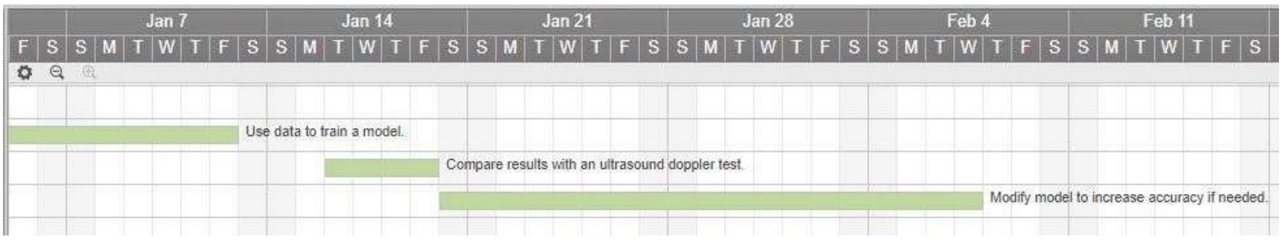


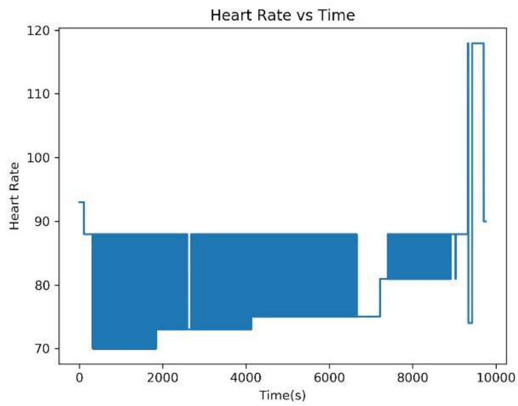
Fig 4.17 Part 5

Chapter 5
Implementation Details

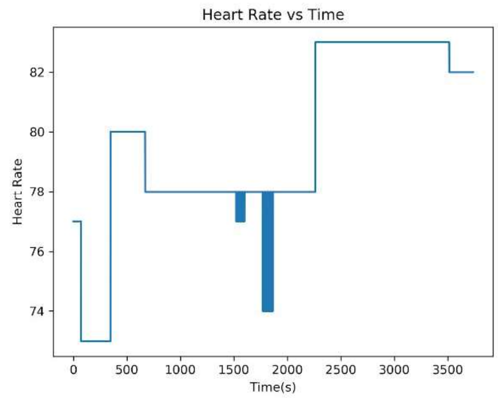
5.1. Data Processing

We are getting data from the Surdial 55 Plus dialysis machine manufactured by Nipro, Japan. This machine is designed for ease of use and comes with a swivelling touch panel to facilitate input and output of patient health parameters. It features a single needle system with a clamp for ultrafiltration as opposed to a dual needle system. The machine checks for and warns if microbubbles are present within the bloodstream to avoid complications such as a gas embolism. It can generate data on a per session, hour, 15 minutes or second basis. As an add-on, the machines we are using are equipped with a LAN port. This allows them to continuously upload data to a local server which has a MySQL database. This same data has been accessed by us via excel files generated by the server. Dialysis technicians are required to enter the patient ID at start of session so that data can be separated by session and patient.

As mentioned above, data has been provided to us in the form of excel files. These files have been generated on a daily basis. A single excel file contains all the dialysis sessions carried out at that centre that day. It holds the values of all the parameters generated and calculated by the machine. The data of a session is historical, starting at the beginning of a session and updating every second till the end of the session. Owing to the requirement of achieving sufficient clearance, the length of a dialysis session is not constant. It varies from patient to patient and from session to session. Since dealing with excel files manually is cumbersome, we convert them via a python script to the CSV format to allow for easy and fast access and manipulations. Having comma separated values, significantly speeds up data processing since we do not need to deal with the XML file structure setup by Excel for its default xslx file format. The CSV files thus created are processed using another python script. In this script, we load the data on a per file basis into a pandas dataframe on which we perform manipulations. The data is then filtered by session and used to generate graphs similar to those shown in Fig 5.1-5.7. These figures have two parts



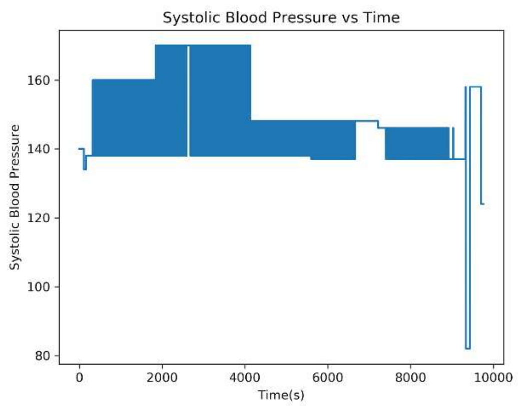
(a)Abnormal



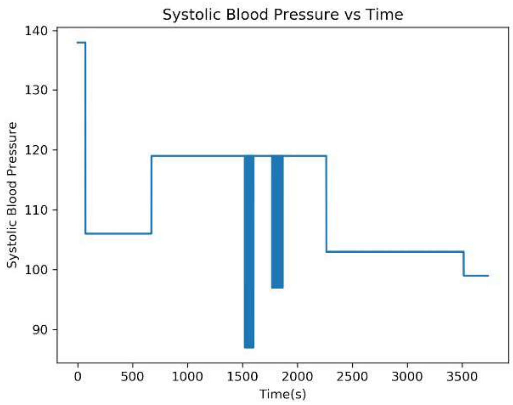
(b)Normal

Fig 5.1 Graphs generated - Heart Rate

Fig 5.1 shows heart rate readings for two patients. It can be observed that readings from abnormal fistula showcase a similar heart rate trend with a rise towards the end.



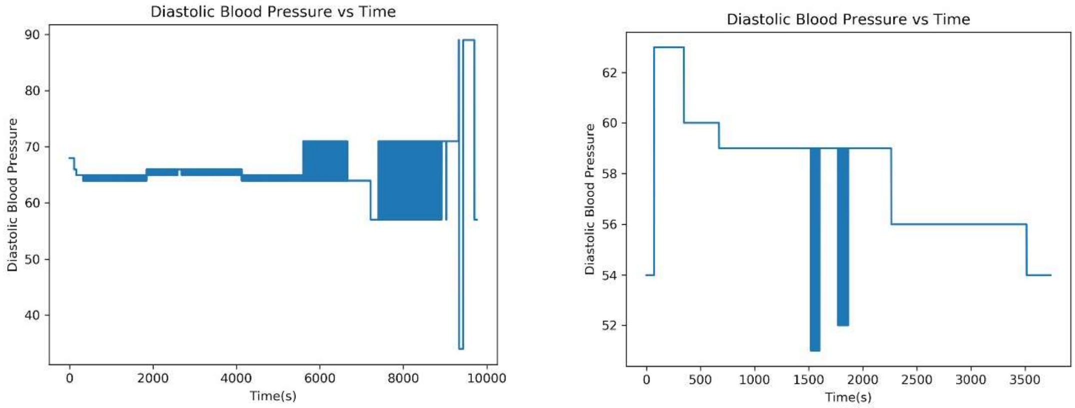
(a)Abnormal



(b)Normal

Fig 5.2 Graphs generated - Systolic Blood Pressure

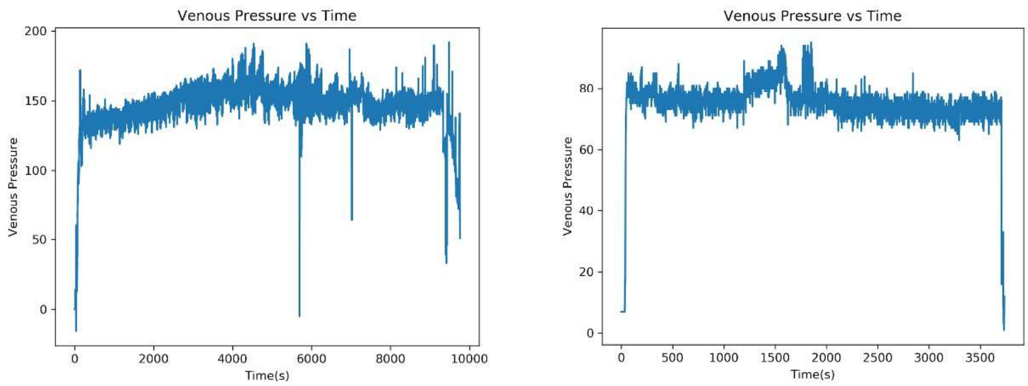
Fig 5.2 shows the systolic blood pressure readings for two patients. It can be observed that readings from abnormal fistula condition showcase a higher systolic blood pressure.



(a)Abnormal (b)Normal

Fig 5.3 Graphs generated - Diastolic Blood Pressure

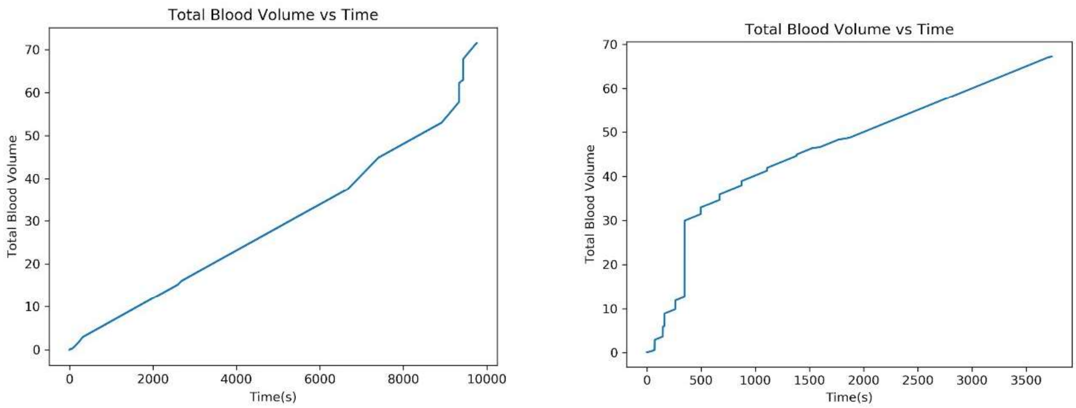
Fig 5.3 shows the diastolic blood pressure readings for two patients. It can be observed that readings from abnormal fistula have less variation in values over time and also have higher values.



(a)Abnormal (b)Normal

Fig 5.4 Graphs generated - Venous Pressure

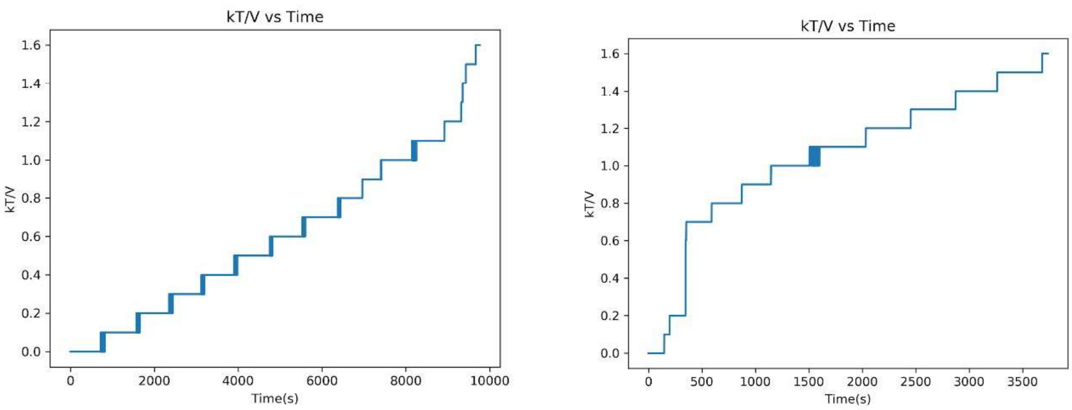
Fig 5.4 shows the venous pressure readings for two patients. It can be observed that readings from abnormal fistula condition are significantly higher compared to those from a good fistula condition.



(a)Abnormal (b)Normal

Fig 5.5 Graphs generated - Total Blood Volume

Fig 5.5 shows the Total Blood Volume readings for two patients. It can be observed that readings from abnormal fistula condition is taking much more time to reach the target treated blood volume as compared to the normal fistula.



(a)Abnormal (b)Normal

Fig 5.6 Graphs generated - kT/V

(k – dialyzer clearance of urea. T – dialysis time. V – volume of distribution of urea, approximately equal to patient's total body water.)

Fig 5.6 shows the kT/V readings for two patients. It can be observed that readings from abnormal fistula condition is taking much more time to reach the target kT/V value (1.2) as compared to the normal fistula.

These graphs are analysed alongside the known health of arteriovenous fistula as determined by the gold standard ultrasound doppler test in order to determine factors relevant to its health. This analysis and discussion with nephrologists led us to identification of parameters to be used. The parameters considered are:

- Heart rate
- Systolic BP
- Diastolic BP
- Venous Pressure
- Arterial Pressure
- kT/v slope
- Cumulative blood volume slope

These parameters are further divided into sub parameters recording the maximum, minimum, mean and standard deviation. The data considered for this analysis has been gathered over a period of one month from the Chembur branch of Apex Kidney Care. After cleaning of data, we have 313 sessions.

5.2. Algorithms for the respective modules developed

Initially, we looked into doing a time series analysis of the data since we had access to historical data over a session. However, since there was quite some variance in the duration of a session, we decided against doing the same. We thus looked into multiple other algorithms to work with our data and decided on the following:

- Support Vector Machines
- Random Forest Classifier
- Multi Layer Perceptron

SVM or Support Vector Machines are a supervised learning method allowing us to do classification and regression analysis on a dataset. It helps to capture more complex relationships between data points without performing any difficult transformations. The kernel takes the data and transforms it which is quite similar to unraveling a strand of DNA. This kernel trick makes SVM more efficient. SVM is able to compute optical hyperplanes and is robust due to optimal margin gap between these separating hyperplanes. We have used a radial basis function kernel for the SVM with a kernel coefficient of 1/n and a tolerance of 0.001 where n is the number of features provided to said algorithm. The algorithm runs till the required tolerances are achieved. There is no hard set limit on the number of iterations to be performed.

The classifier used for implementing SVM is:

svm.SVC(C=1.5, kernel='rbf', degree=3, gamma='auto', coef0=0.0, shrinking=True, probability=False,tol=0.001,cache_size=200,class_weight=None,verbose=False,max_iter=-1, decision_function_shape='ovr', random_state=None)

Random Forests is a supervised ensemble classification algorithm where ensemble classifiers are randomly created decision trees. These trees acts as classifiers and the target prediction is based on the majority voting method. Random forests require no input preparation as they are capable of handling binary features, categorical features, numerical features without scaling. The Random Forest algorithm can be used for identifying the most important features from the training dataset, in other words, feature engineering. They can also be easily grown parallely unlike boosted models or large neural networks. In our use case, the algorithm fits 50 decision trees on sub samples of the data set which are drawn with replacement and are of the same size as the initial dataset, the number of jobs is set to number of cores to minimise training time and all attributes are utilized while forming a decision tree. This classifier uses averaging to improve accuracy and have a control on overfitting.

The classifier used for implementing Random Forest is:

RandomForestClassifier(n_estimators=50,criterion='gini',max_depth=None,min_samples_split=2,min_samples_leaf=1,min_weight_fraction_leaf=0.0,max_features='auto',max_leaf_nodes=None,min_impurity_decrease=0.0,min_impurity_split=None,bootstrap=True,oob_score=False,n_jobs=-1,random_state=1,verbose=0,warm_start=False,class_weight=None)

Neural networks have an ability to derive meaning from complicated and/or imprecise data. MLP or Multi-Layer Perceptron is a supervised learning algorithm which is a form of feedforward artificial neural network with at least three layers of nodes. Every neuron except for the input nodes uses a non linear activation function. It trains using back propagation and is useful for non linear patterns. Generally, the stochastic gradient descent function is used for the log-loss function in this algorithm but we have opted to use the low memory BFGS function instead since it converges faster and is reported to have better performance when it comes to smaller data sets.

A MLP with hidden layers of sufficient size can approximate any continuous function to any desired accuracy. Unlike other prediction techniques, MLP does not impose any restrictions on the input variables. Studies have shown that MLPs can better model heteroskedasticity i.e. data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data.

The classifier used for implementing Multilayer Perceptron is:

MLPClassifier(solver='lbfgs', alpha=1e-5, hidden_layer_sizes=(15,), random_state=1)

We chose these algorithms since they can tolerate smaller datasets and since practical usage of the same gave a high accuracy.

The basic steps used for implementing the above mentioned algorithms are:

1. Read the CSV input.
2. Divide the dataset into training (80%) and testing (20%) data.
3. Train the training data using its respective classifier.
4. Test the data
5. Calculate
    a. False positive accuracy
    b. False negative accuracy
    c. True positive accuracy
    d. True negative accuracy
    e. Overall accuracy
6. Generate graphs using the above mentioned accuracies

5.3. Comparative Analysis of algorithms

Initially, we trained all three algorithms on the dataset. However, there was a major imbalance between data of bad fistulas and good ones. Good fistulas outnumbered bad ones 85:15. To reduce this imbalance, we introduce two cases: case 1 and case 2 respectively.

In case 1, we have 313 sessions out of which 42 are bad fistulas and 271 are good fistulas. We have done a 80-20 split of data for training and testing to maintain a blind test.

Algorithm	Case 1			Case 2		
	Accuracy	Prediction - Good	Prediction - Bad	Accuracy	Prediction - Good	Prediction - Bad
SVM	88%	98%	0%	87%	95%	25%
RFC	96%	98%	0%	91%	94%	86%
MLP	92%	100%	0%	91%	88%	100%

Table 5.1 Comparison of accuracy

In case 2, we have 111 sessions out of which 42 are bad fistula and 69 are good fistulas. In this case, we are reducing the imbalance to reduce the overfitting. We have done a 80-20 split of data for training and testing to maintain a blind test.

Table 5.1 shows the accuracies of all three algorithms over both cases.

We find that, while the accuracy in the case of Case 1 seems acceptable, the prediction of bad fistula is wrong every single time. The reason for this is the imbalance between the number of sessions for good and bad fistula. The imbalance between the two is likely to reduce over time as the number of dialysis centres providing data increases. In this case, we are prone to underfitting since the data is sparse for bad fistulas. In Case 2, we have reduced the imbalance in the dataset and are due to this the prediction rate for bad fistulas rose. In the case of the multi-layer perceptron, it actually reached 100% while that of good fistula dropped to 88%. For the required use case, identifying bad fistulas for checking, it is much better to identify a good fistula as bad rather than a bad fistula as good, endangering a patient's life. Hence we are deploying the MLP model for our desktop application.

Chapter 6

Testing

Every time the model is trained, we have done a 80:20 split of dataset into separate datasets for training and testing in order to ensure that we have a blind test and that the system is not tested with the same data it was trained with. Ahead of the same, in our training module, if we want to update the model, we run a check to see whether the new model gives us better results than the current one. Only if it does do we replace the save file of the model with the updated one. This serves to stop from degrading the performance of our system. For ex, in the fig. 6.1, the initial model formed has an accuracy of 77%. The model formed with 71% is rejected as it's accuracy is less than the previous one whereas the model with accuracy is accepted and stored. The new model replaces the previous model and the process continues unless and until the process reaches its termination value.

```
Iteration count: 0
accuracy = 0.7792207792207793
Iteration count:  1
Accuracy:  0.7142857142857143
The model is discarded
Iteration count:  2
Accuracy:  0.8246753246753247
The model is accepted and hence stored
```

Fig. 6.1 Iterations for accepting model

The model is stored in a file within the same folder or to a particular path in a device. For ex, the model of Multi-layer perceptron (Fig 6.2) is being stored in finalized_model_mlp.sav. The same file will be used later for prediction as well.

finalized_model_mlp.sav 19-04-2018 01:00 SAV File 6 KB

Fig. 6.2 Saved Model

Chapter 7
Result Analysis

7.1 Simulation Model

The application begins with a login module which supports 2 different kinds of users. Doctors and Support staff. The admin profile will be used by the tech team. For the Doctor profile there are 5 functionalities available-

- Analysis
- Prediction with CSV input
- Prediction with manual input
- Training and testing
- Patient data
- Graphical analysis

The Analysis module presents overall accuracy of 3 algorithms used and the trend in accuracy over varying dataset sizes.

Prediction with CSV input module allows the user to predict the health of multiple patients using a single CSV file with proper values for each parameter.

Prediction with manual input works in the same way for a single patient but the value of each parameter has to be entered manually in the text fields.

Training and testing module allows the user to train the algorithms with new or updated data sets.

History of all patients checked and their report is displayed in the Patient data page.

Graphical analysis module presents a comparative graphical analysis of the 3 algorithms-RFC,MLP,SVM with the accuracy of good,bad fistulae and the overall accuracy of the same.

The support staff account is more restrictive with only the Prediction with CSV input, Prediction with manual input modules present. Also each support staff can see the past tests done by him and the corresponding result.

7.2 Parameters considered

The graphs we generated were analysed alongside the known health of arteriovenous fistula as determined by the gold standard ultrasound doppler test in order to determine factors relevant to its health. This analysis and discussion with nephrologists led us to identification of parameters to be used. The parameters considered are:

- Heart rate
- Systolic BP
- Diastolic BP
- Venous Pressure
- Arterial Pressure
- kT/v slope
- Cumulative blood volume slope

These parameters are further divided into sub parameters recording the maximum, minimum, mean and standard deviation.

7.3 Screenshots of User Interface (UI)

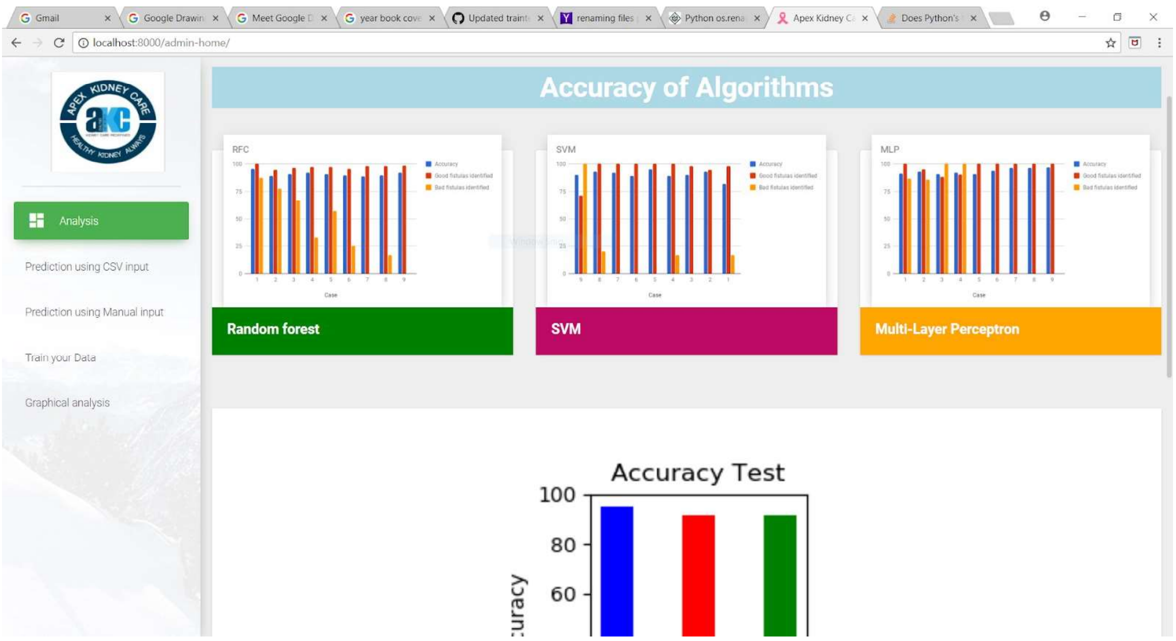


Fig 7.1 Analysis Page

The admin landing page of the system i.e. fig 7.1 is Analysis page which contains the overall dynamic accuracy graph for algorithms i.e. Support vector machines, Random forest algorithm, Multilayer perceptron model. This place also includes graphs related to variation in the the datasets on different models, blue signifies total accuracy, red is for true positives and orange signifies true negatives.

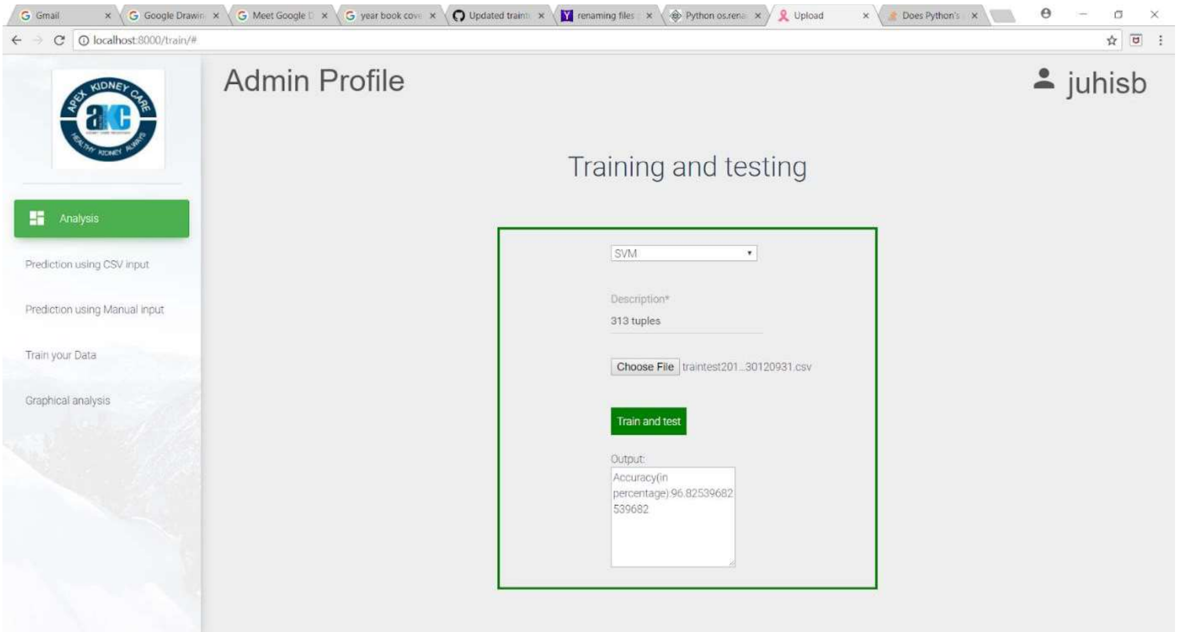


Fig 7.2 Training and testing module

Fig 7.2 is the screenshot of Training and testing module of the system. The dropdown is provided to choose from the given list of the algorithms for training. Description field is provided to input information about the dataset on which the model will be trained. User needs to upload the csv file for training and testing. Results are displayed in terms of accuracy of the current trained model of chosen algorithm. This functionality is provided to admin profile only.

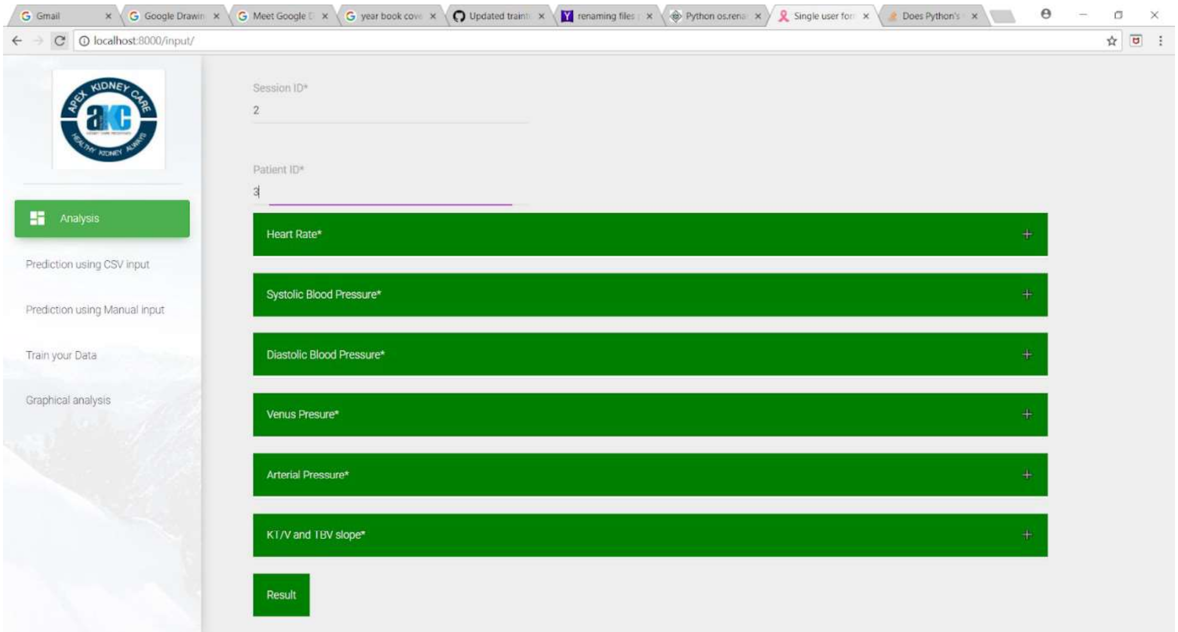


Fig 7.3 Prediction using single input

Module in the fig 7.3 is Prediction module in which input values are manual input from the user. Input fields are provided for each parameter required they are Session ID, patient ID, heart rate, systolic blood pressure, diastolic blood pressure, venous pressure, arterial pressure, total blood volume slope, KT/V slope. On submitting these inputs a downloadable report with prediction result is generated.

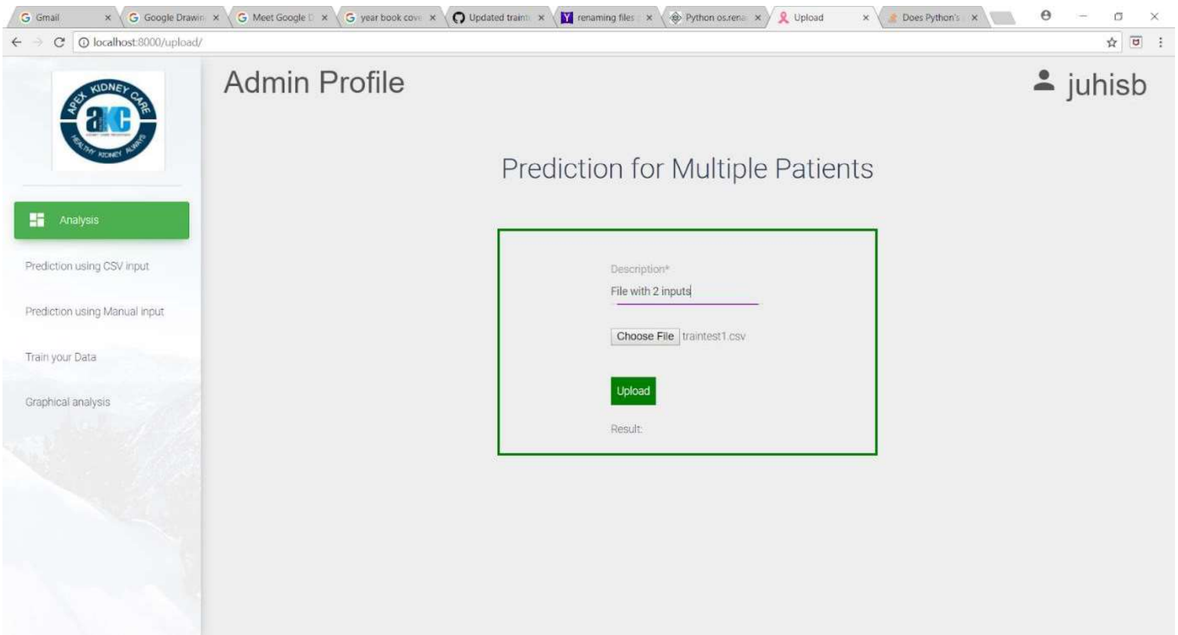


Fig 7.4 Prediction using csv input

Module in the fig 7.4 is Prediction module in which input values are taken from the CSV file. The CSV file contains the values for Session ID, patient ID, heart rate, systolic blood pressure, diastolic blood pressure, venous pressure, arterial pressure, total blood volume slope, KT/V slope for multiple patients. On submitting this files a downloadable report with prediction result for all the patients is generated.

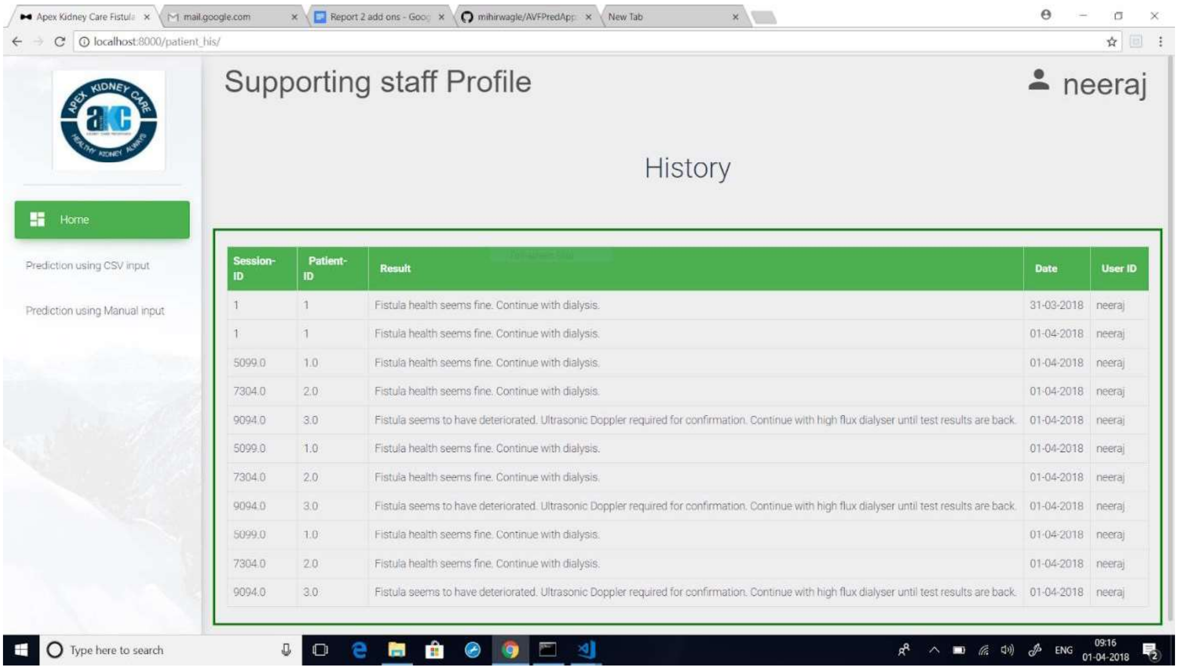


Fig 7.5 Patient history

Patient history is displayed with the parameters such as Session-ID of the dialysis, Patient-ID, Results of the dialysis session , date and User ID which is the id of the staff who performed the prediction as shown in the fig 7.5 which is the screenshot of the module.

7.4 Graphical outputs

The graphs shown in this section are dynamic graphs and depends on the CSV file used for the training of the prediction module. According to our code, input data gets divided into two parts: training data (80%) and testing data (20%). Information is fetched from the CSV input which tells us about the percentage of bad fistula and good fistula present in the data. It is followed by the accuracies of false positives, false negatives and the overall accuracy of each algorithm. It helps us to get a proper insights of the algorithms used.

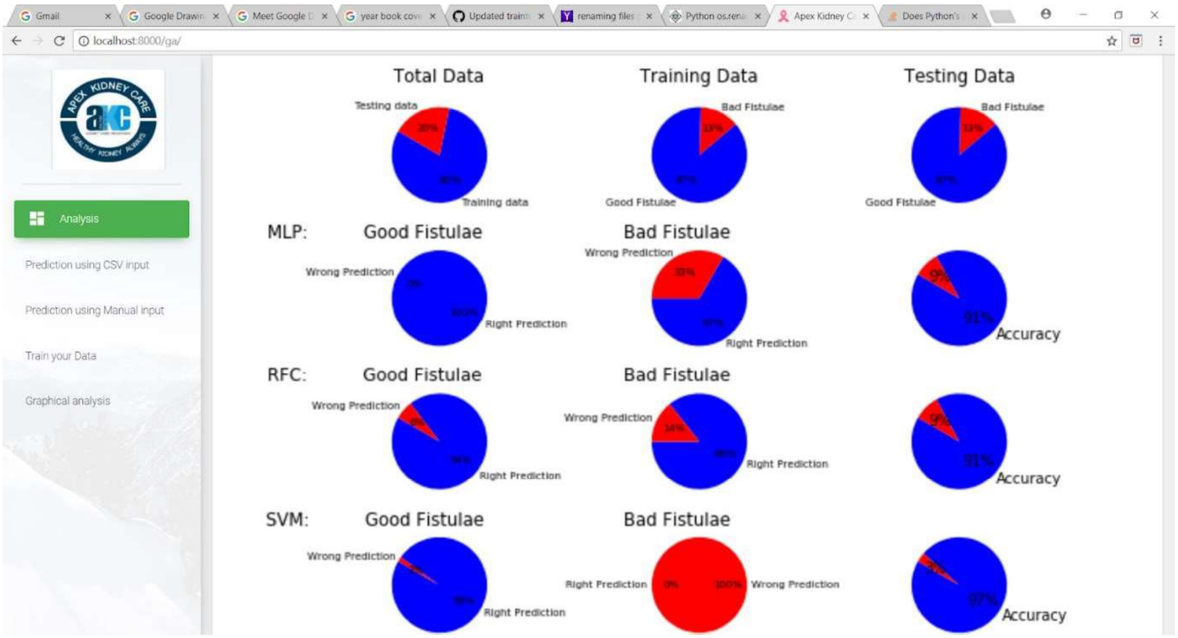


Fig 7.6 Graphical Analysis Module

The graphs given below are formed considering nine cases of dataset. The dataset considered for the prediction model kept increasing from case 1 to case 9. Case 1 consisted of 110 tuples and Case 9 consists of approx. 313 tuples. The main purpose of this activity was to find out the efficiency of different algorithms in different scenarios.

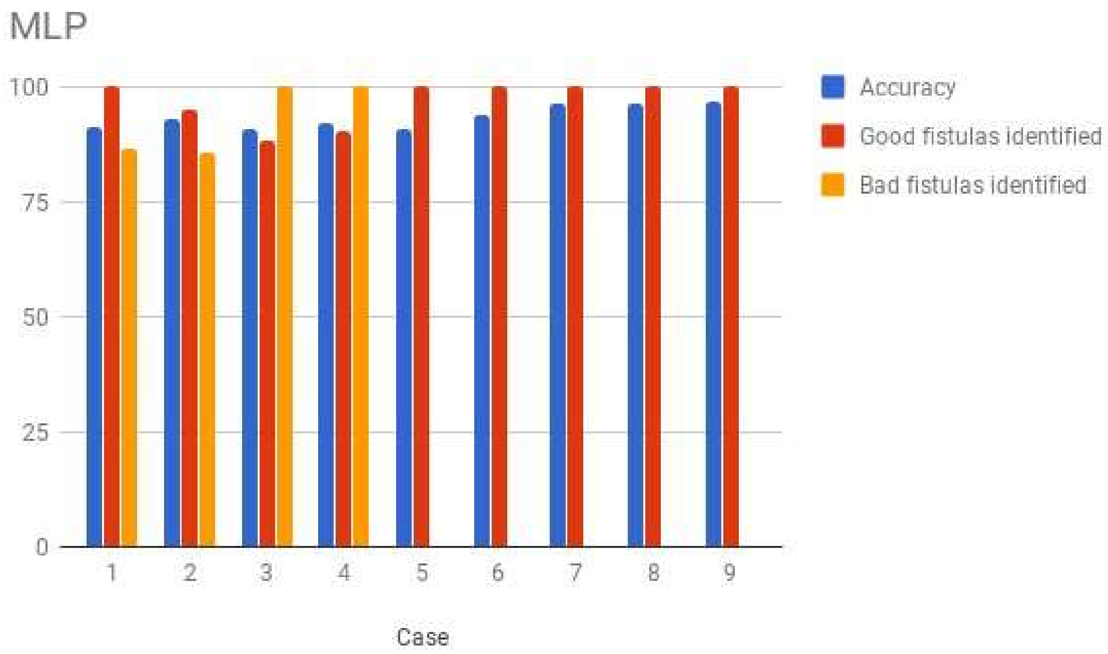


Fig 7.7 Accuracy of MLP over varying dataset size in percentage

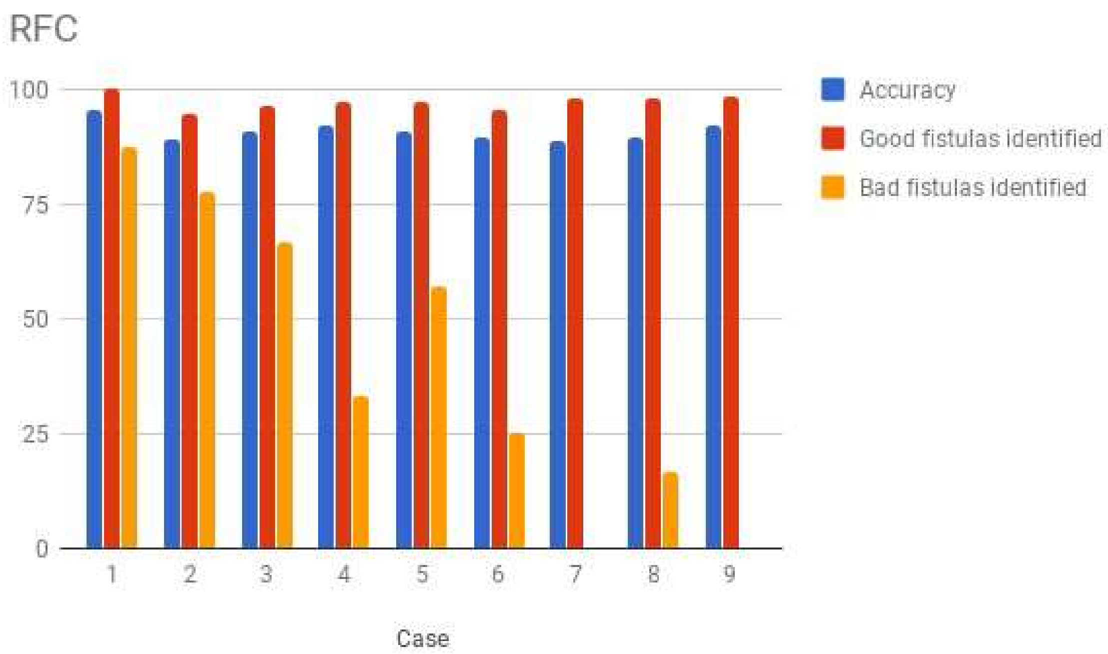


Fig 7.8 Accuracy of RFC over varying dataset size in percentage

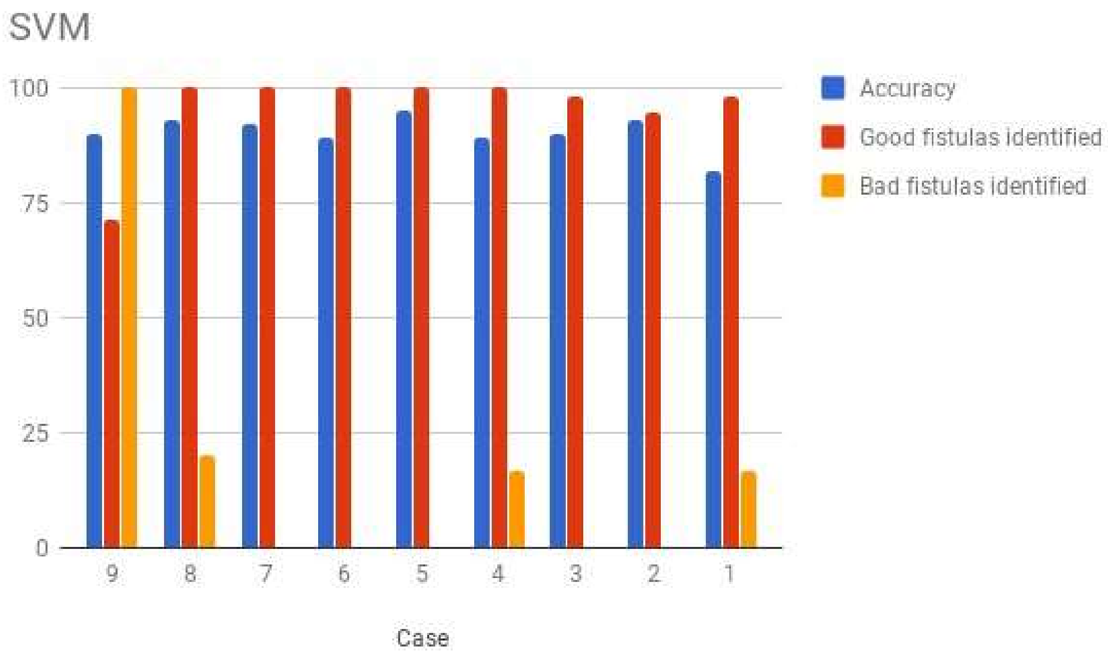


Fig 7.9 Accuracy of SVM over varying dataset size in percentage

7.5 Reports generated / Tables obtained

The report generated by the prediction model consists of all the factors considered for the prediction of the state of the arteriovenous fistula. It informs about the state of the fistula and suggests for Ultrasound Doppler incase of an unhealthy fistula. Fig 7.10 shows a sample report.

Patient Report 1

Session ID: 2
Patient ID: 1
Heart Rate --
Max Heart Rate: 90.0
Min Heart Rate: 67.0
Max-Min Heart Rate: 23.0
Mean Heart Rate: 71.89651335
Std Heart Rate: 4.533112667
Systolic Blood Pressure --
Max Systolic Blood Pressure: 206.0
Min Systolic Blood Pressure: 169.0
Max-Min Systolic Blood Pressure: 37.0
Mean Systolic Blood Pressure: 178.29209939999998
Std Systolic Blood Pressure: 8.118933028999999
Diastolic Blood Presure ---
Max Diastolic Blood Presure: 115.0
Min Diastolic Blood Presure: 86.0
Max-Min Diastolic Blood Presure: 29.0
Mean Diastolic Blood Presure: 96.45252226
Std Diastolic Blood Presure: 8.542753627
Venous Pressure---
Max Venous Pressure: 150.0
Min Venous Pressure: -70.0
Max-Min Venous Pressure:
Mean Venous Pressure: 112.2156899
Std Venous Pressure: 16.6033307
Arterial Pressure ---
Max Arterial Pressure: 1.0
Min Arterial Pressure: 0.0
Max-Min Arterial Pressure: 1.0
Mean Arterial Pressure: 0.511683976
Std Arterial Pressure: 0.499909825
Kt/V --
Kt/V Slope: 0.00033382800000000004
TBV ---
TBV Slope: 0.01263724
Result --
Prediction: Fistula seems to have deteriorated. Ultrasonic Doppler required for confirmation. Continue with high flux dialyser until test results are back.

Fig 7.10 Sample Report

Chapter 8
Conclusion

8.1 Limitations

The system we have created uses data directly from the dialysis machines. In this case, it is only one type of dialysis machine Nipro Surdial 55 Plus. However not all dialysis machines can provide the exact same data points. Also, the data is limited to 310 sessions due to the same reason.

8.2 Conclusion

This project proposes the use of technology in predicting the health of an arteriovenous fistula. Over time, the health of the fistula is likely to worsen and eventually fail. If this happens without warning, the patient is inconvenienced by catheter dialysis until a new fistula can develop. One way to get around this is by carrying out the ultrasound doppler test on a monthly basis. However this test is expensive and places a burden on the patient and also takes up the doctors' time. The system we have developed uses data received directly from the dialysis machine and leverages it to predict health of fistula. This check is done each time dialysis is carried out. If it throws up a red flag, we direct the patient to carry out the ultrasound doppler test to confirm the diagnosis. Thus the cost to the patient is minimised and doctors can create a new fistula in time to avoid inconveniencing the patient with catheter dialysis. Thus our system is likely to improve quality of life of the patient as well as reduce their medical costs. It will also save the doctors' time.

8.3 Future Scope

The accuracy of the system using the Multilayer perceptron model can be improved by studying the correlation between the number of hidden layers and the number of parameters considered for prediction to find the optimal number of hidden layers. We plan to carry out feature selection using PCA(Principal Component Analysis) to avoid considering parameters that may be correlated. The training data will be updated with data of more patients and also try to include more tuples with bad fistulas as the prediction class. Also, different brand machines may give data to their brand specific database. All this data has to be converted and dumped into a single warehouse where we can run triggers to generate our dataset. Doing the same will allow us to expand our data set significantly and therefore also the accuracy and efficiency of the system.

Chapter 9

References

[1] Norbert Lameire, Jean-Pierre Wauters, Jose Luis Górriz Teruel, Wim Van Biesen, Raymond Vanholder, 'An update on the referral pattern of patients with end-stage renal disease', Kidney International Vol. 61, pages S27-S34, May 2002.

[2] Homer William Smith, 'The kidney: structure and function in health and disease', Oxford University Press, USA Vol. 1, 1951.

[3] Arif Asif, Prabir Roy-Chaudhury, Gerald A Beathard, 'Early arteriovenous fistula failure: a logical proposal for when and how to intervene', Clinical Journal of the American Society of Nephrology Vol. 1, issue 2, pages 332-339, March 2006.

[4] Rajiv Agarwal, Allen R Nissenson, Daniel Batlle, Daniel W Coyne, J Richard Trout, David G Warnock, 'Prevalence, treatment, and control of hypertension in chronic hemodialysis patients in the United States', The American journal of medicine Vol. 115, issue 4, pages 291-297, Sept. 2009.

[5] Pasquale Zamboli, Fulvio Fiorini, Alessandro D'Amelio, Pasquale Fatuzzo, Antonio Granata, 'Color Doppler ultrasound and arteriovenous fistulas for hemodialysis', Springer International Publishing Vol. 17, Issue 4, pages 253-263, Dec. 2014.

[6] Manoj Reddy and John Cho, 'Detecting Chronic Kidney Disease Using Machine Learning', Qatar Foundation Annual Research Conference Proceedings Vol. 2016, issue 1, March, 2016.

[7] BV Ravindra, N Sriraam, M Geetha, 'Discovery of significant parameters in kidney dialysis data sets by K-means algorithm', IEEE, pages 452-454, Nov. 2014.

[8] Jeffrey Navarro Rojas, 'A Systematic Review on the Quality of Life among Hemodialysis Patients in the Middle East Countries', International Journal of Science and Research Vol. 6, issue 5, pages 532-542, May 2017.

[9] Enes Celik, Muhammet Atalay, Adil Kondiloglu, 'The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning', International Journal of Intelligent Systems and Applications in Engineering Vol. 4, issue special issue-1, pages 27-31, Dec. 2016.

[10] Aaron Stern, Soumya Sachdeva, Rohit Kapoor, Jasjit Singh, Sarthak Sachdeva, 'High blood pressure in dialysis patients: cause, pathophysiology, influence on morbidity, mortality and

management', Journal of clinical and diagnostic research: JCDR Vol. 8, issue 6,pages ME01, 20 Jun. 2014.
[11] Jeremy Levy, Edwina Brown, Anastasia Lawrence, 'Oxford Handbook of Dialysis', Oxford University Press, Feb. 2016.
[12] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, 'Tensorflow: Large-scale machine learning on heterogeneous distributed systems', arXiv preprint arXiv:1603.04467, March 2016.
[13] Spyros Xanthopoulos, Stelios Xinogalos, 'A comparative analysis of cross-platform development approaches for mobile applications', ACM, pages 213-220, Sept. 2013.
[14] Congcong Li, Jie Wang, Lei Wang, Luanyun Hu, Peng Gong, 'Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery', Remote Sensing Vol. 6, issue 2, pages 964-983, Jan.
2014.

Chapter 10

Project Review Sheets

10.1 Project review sheet 1

Company:- Apex kidney care.
contact person :- Vishwanath Billa - MD, DM (Nephrology), Director (AKC) 9820140041

Project Evaluation Sheet 2017 - 18 GIB

Class: D17 ~~A~~/B/~~C~~
Group No.: 06

Title of Project: Application of Machine learning techniques for the analysis and prediction of hypertension and vein function in hemodialysis.

Group Members: Juhi Bhagtani, Aishwarya Chandak, Neeraj Jethnani, Mihir Wagle.

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	5	3	5	1	5	2	2	2	1	3	2	3	5	5	44

Comments: Need to reconstruct title of project. Intrested to see test cases (real-time) and final implimentation on m/c.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	3	5	5	2	5	2	2	2	1	3	1	3	5	4	43

Comments: Topic / title of project reconstruct. Collecting test cases and actual implimentation is challenging.

Date: 26th September 2017

Snehal More
Name & Signature Reviewer2

10.2 Project review sheet 2-1

Project Evaluation Sheet 2017 - 18

Title of Project: Application of Machine learning techniques for the analysis of hypertension and prediction of vein function in hemodialysis. GROUP NO.: 6

Group Members: Aishwarya Chandak, Juhi Bhagtani, Neeraj Jethnani, Mihir Wagle.

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 2	4	4	3	2	2	2	2	2	2	1	1	2	4	3	35

Comments: Video proven on complete working + doctor's signature + solution to patients.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 2	4	3	4	3	2	2	2	2	2	2	2	1	3	3	35

Comments: Presentation need improve. Verification from doctore necessary

Date: 26/02/2018

Snehal Mane 26/2/18

Name & Signature Reviewer2

10.3 Project review sheet 2-2

Inhouse/ Industry: Industry project.

Class: D17 ~~A~~/B/~~C~~

Group No.: 6

Project Evaluation Sheet 2017 - 18

Title of Project: Application of Machine learning techniques for analysis and prediction of hypertension and vein function in hemodialysis.

Group Members: Aishwarya Chandak, Juhi Bhagtani, Neeraj Jethnani, Mihir Wagle.

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2	1	2	2	3	5	5	41

Comments: All changes done.

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2	1	2	2	3	5	5	41

Comments: Very Good.

Date: 15th March,2018

15/3/18 Snehal Mane

Name & Signature Reviewer2

Chapter 11

Appendix

11.1 List of Figures

11.2 List of Tables

2018 3rd International Conference for Convergence in Technology (I2CT)
The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018

Machine Learning for Prediction of Life of Arteriovenous Fistula

Dr. Gresha Bhatia, *Deputy HOD, CMPN*, *V.E.S.I.T*, Mumbai, India
Mihir Wagle, *Computer Engg*, *V.E.S.I.T*, Mumbai, India
Neeraj Jethnani, *Computer Engg*, *V.E.S.I.T*, Mumbai, India
Juhi Bhagtani, *Computer Engg*, *V.E.S.I.T*, Mumbai, India
Aishwarya Chandak, *Computer Engg*, *V.E.S.I.T*, Mumbai, India

***Abstract*—Millions of patients worldwide suffer from Kidney failure and require dialysis. In most cases, dialysis is started after the kidney function of the patient falls below a threshold. In this scenario the patients kidney is essentially non functional. In order to conduct dialysis, native arteriovenous fistulas are constructed to increase blood flow in the superficial vein, and hence facilitate dialysis. Over time, as dialysis continues, the patient may suffer from hypertension and reduced vein function leading to the collapse of the fistula. The ultrasound doppler test for checking the state of the fistula are expensive and doing it again and again is not feasible. We study work related to Chronic Kidney Disease which provides similar data points to those required for the health of the fistula and propose a mechanism to predict the life of a fistula.**

***Index Terms*—Dialysis, Machine Learning, Fistula, Chronic Kidney Disease**

I. Introduction

Kidney failure or renal insufficiency, is a condition which exhibits damaged kidney function where the kidneys lose their ability to excrete metabolic wastes from the blood stream [1]. The two main types include acute kidney injury(AKI), which may be reversed by timely remedies, and chronic kidney disease, which is unfortunately often irreversible.

Chronic kidney disease (CKD) is a gradual decrease in renal function over an extended period, sometimes months, generally years. When a patient is diagnosed as having CKD, he needs to be put on dialysis, for which a native arteriovenous fistula needs to be constructed [1], [2]. Dialysis is generally done for four hours three times a week to ensure acceptable Quality of life for the patient. Over time, as dialysis is conducted, some patients might suffer from hypertension and reduced vein function which may eventually lead to the fistula failing [3], [4]. Developing another fistula takes time and during this period, the patient is inconvenienced by dialysis via a catheter [3]. Constant ultrasound doppler scans[5] to check fistula health are not feasible. Currently, the very significant amount of data being output by the dialysis machines is view and go. I.e. It is not stored. It is not used except during dialysis. Therefore, in the current system, there is no prediction and most procedures follow a action response method where a new fistula is created once the old one fails catastrophically and hypertension is not treated proactively [5].

The focus of this paper is to utilise this data in order to predict fistula health, streamlining procedures carried out by nephrologists, allowing them to proactively treat patients to stop them from entering a hypertensive state and keeping them informed as to the state of the fistula, to allow for the creation of a new one before the current one fails so as to not inconvenience a patient with a catheter as a stopgap measure. Doing the same is likely to improve the quality of life of the patient allowing them to lead a mostly normal life even while they undergo dialysis. This paper is further divided into five more sections. Section II provides an overview of renal disease covering its types and types of dialysis possible. Section III discusses the work already done in this field and all our inferences. Section IV covers the work proposed, including all the modules to be included. Section V covers the modules and a brief description of each. Section VI shows a flow diagram for the work proposed. Section VII mentions evaluation details. Section VIII discusses the work done thus far. Section IX concludes the paper by discussing the problem at hand, its solution and implications.

II. Renal disease - An Overview

Patients suffering from renal impairment, generally present themselves with symptoms which seem to be mostly non specific or if a consultation with a medical professional has revealed elevated urea or creatinine. When such a patient consults a nephrologist, it is of paramount importance that he or she is able to identify the ailment of the patient and distinguish between AKI and ESKD [2].

AKI stands for Acute Kidney Injury. It is generally a sudden onset episode of kidney failure or damage which results in a severe imbalance of body fluids due to a build up of excretory products in the blood stream. It may also be caused by reduced blood flow to the kidneys or a blockage in the urinary tract. In most cases, this disease is treatable and kidney function may be recovered. ESKD stands for End Stage Kidney Disease. As the name suggests, this is a fatal disease and reversing the damage caused is no longer possible[6]. A patient can be said to have ESKD when his or her kidney function falls below 15%. This can be considered the last stage of Chronic Kidney Disease(CKD) and severely shortens life expectancy[2]. At this stage, it is almost certain that the patient has lost his kidneys. The patientis now

978-1-5386-4273-3/18/$31.00 ©2018 IEEE

presented with two options:

- Lifelong dialysis.
- Dialysis followed by Kidney transplant.

Patients and their relatives can add themselves to transplant list to enroll themselves for a domino transplant or relatives can donate a kidney if there is a match. While on the list, dialysis occurs regularly. In case the patient gets a transplant, it can be stopped, else it continues for life When it comes to undergoing dialysis, the patient has four options:

- Temporary catheter.
- Distal arteriovenous fistula.
- Proximal arteriovenous fistula.
- Permanent catheter.

A temporary catheter is a stopgap measure which allows a patient to undergo dialysis for the two weeks it takes for a fistula to develop. Of the two techniques for the arteriovenous fistula, the proximal method is preferred over the distal method. In the distal method the fistula is constructed near the wrist. The proximal method constructs it in the upper arm. If a fistula is no longer possible, a permanent catheter may be used. However, this inconveniences the patient and has a short lifespan of around one and a half year compared to fifteen to twenty years for a fistula.

III. Related Work

A lot of work has already been done in the field of dialysis on the whole. Much of the work done concerns the various methods and techniques involved in identifying Chronic Kidney Disease.
Paper[7] deals with the identification of Chronic Kidney Disease using machine learning techniques. This paper suggests trends in algorithms used. It does not specify the parameters considered and states that the training set did not have 100% observed parameter values. It specifies 6 different classification algorithms that were used to compare them. They include: logistic regression, decision tree, SVM with a linear kernel, SVM with a RBF kernel, Random Forest Classifier and Adaboost. Of these, SVM with linear kernel gives the highest accuracy of 98 percent.
Paper[8] looks to achieve the same goal but using Decision trees and SVM. In this scenario, Decision trees give an accuracy of 97%-100% whereas SVM gives an accuracy of 97%. The results are based on the population of 250 patients with CKD and 150 healthy patients. Sequential minimal optimization and J48 was used for decision tree using WEKA and the dataset considered had 25 distinct parameters.
Paper[9] tries to find significant parameters in kidney dialysis sets using the K-means algorithm. It relies on classifying parameters into ranges such as medium, low and high to further aid with clustering. This paper mainly focused on identifying survival period of patient undergoing dialysis using clustering techniques. Making classes helps with making the prediction. In this scenario, creatinine plays an important part and it is found that patients with a level of creatinine which is either high or low are likely to suffer from adverse effects.
Paper[10] deals with the effects of dialysis to the quality of life of a patient. It shows the results of a survey of patients undergoing hemodialysis. Starting hemodialysis involves a significant lifestyle change and can have a lot of effects on a patients physical and mental health. Care needs to be taken that changes are not averse. The paper was restricted to specific region and results cannot be generalized.
Interactive sessions with nephrologists have furthered our understanding of the dialysis process and its subprocesses. These interactions have also helped us come up with relevant factors to be considered in relation to health of the fistula.
Looking at the manual of the Fresenius 4008s, a common dialysis machine, we see that it shows a variety of factors on its display, updating second on second. Most of this data is considered in the present and not stored.
Looking at all the related work it was identified that identification and classification is an important aspect of work in the field of dialysis. However there is a need to provide for forecasting and prediction of life of arteriovenous fistula by making use of state of the art techniques such as Machine Learning and big data.

IV. Proposed Work

This paper presents the various blocks that need to be implemented in a system to predict life of arteriovenous fistulas. The modules in the system are as follows:

- Personal and Clinical Factors
- Analysis of information
- Prediction
- Reports

V. Modules to be developed

The proposed system was envisioned after considering the data set of 200 patients. Firstly, normalcy was defined in the dataset and then the resulting data set was compared. Of the 200 patients 35 had normal functioning of arteriovenous fistula while the remaining had reduced function. 5 dialysis sessions were included for each patient to avoid any case of operator error.
The dataset comprises of over 30 values reported by the machine. Some of them are:

- Dialysis flow
- Blood flow
- Arterial Pressure
- Venous Pressure
- Ultrafiltration speed
- Heart rate
- Trans membrane pressure
- Total Ultrafiltration speed
- Blood pressure

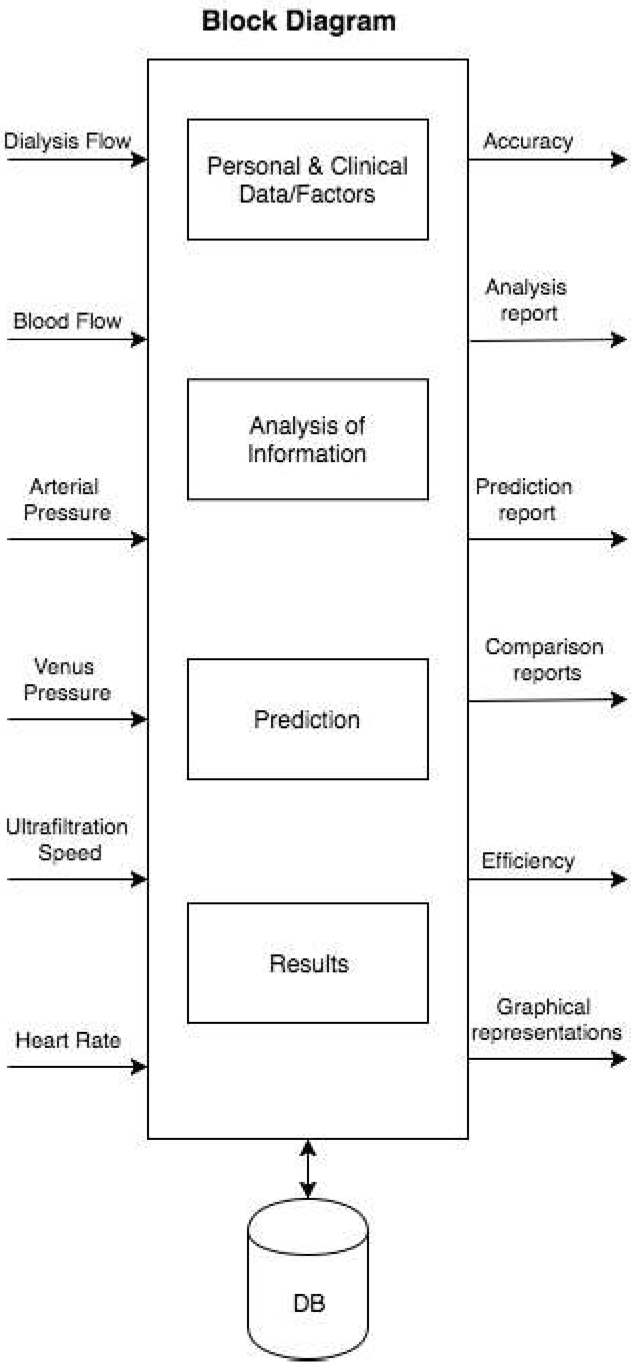


Fig. 1. Block diagram of proposed system

- Target kT/V

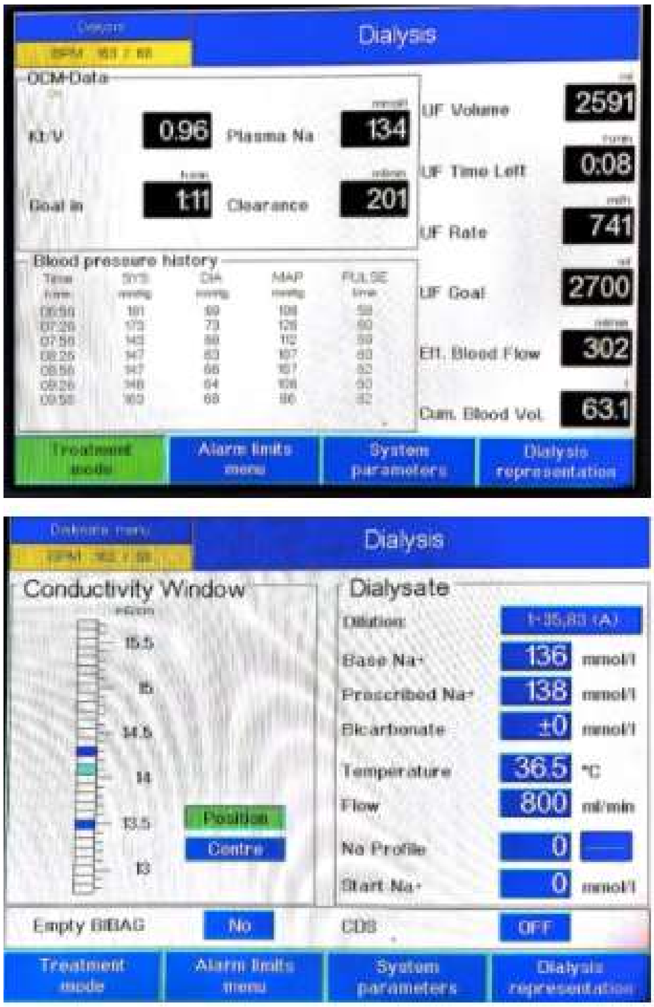


Fig. 2. Fresenius 4008s screengrab

In Fig 2 we see two screen grabs from the Fresenius 4008s dialysis machine. The device has a default treatment mode which is shown in the first image and others for modifying the levels of alarms to ring, system parameters as well as dialysis representation which is shown in the second image.
There are four distinct modules need to be developed in the system in order to effectively predict the life of arteriovenous fistula The first module is Personal Clinical data. This module will take in data which is being output from the dialysis machine and store it within a database after segregating it based on the type of data. Data may be divided into the categories - personal, clinical or miscellaneous.
The second module is Analysis of Information. This module will take in data from the database and process it according to rules to group it into classes by performing segmentation. Alongside segmentation, the data will also be rendered into multiple combinations which will make for clear pattern analysis.
The next module in our system is Prediction. This module will

consider the algorithms we have seen used with hemodialysis and check their efficacy on the data at hand. Based on the efficiency new algorithms may be developed to improve accuracy of the system.
The last module is Results. In this module, we will compare the predicted value with the actual condition to calculate the accuracy of the system and generate reports to track performance of the system.
The hardware required for satisfactory performance of the system consists of a minimum of:

- 2 Intel E5-2670v3 12 Core/24 Thread 2.3Ghz 30Mb Cache
- Processors
- 64 GB RAM
- Nvidia Quadro P600 - 4 units
- Internal drives - 1024GB Samsung 960 Pro NVMe SSD
- External storage - Distributed Ceph Storage with SSD Caching Tier using 800GB Intel SSD DC S3610 Series Drives
- Fresenius 4008s dialysis machine

VI. Implementation of Proposed System

The flow of the system is as follows(Fig 3):

- Technician logs into the system.
- Download of data output by the dialysis machine.
- Upload of csv file to server.
- Shell script to convert csv file to database entries.
- Running script to add to database.
- Add data to training set.
- Use data to train model.
- Test prediction against known patient condition. Optimise algorithms if necessary.
- Generate reports.

Fig 4 shows the the level 1 DFD of the system showing flow of data between the modules.
Once data has been output by the machine, the technician will download it and upload it to the server in csv format.
The data available will be trained in TensorFlow[11] using multiple Machine Learning algorithms including SVM with different kernels, random forest classifiers, logistic regression, adaboost, decision trees and the K-means algorithm [7], [8], [9], [12].
After finding the accuracy for each of these, an algorithm can be developed to increase accuracy of the system. Each parameter will be divided into at least three segments which will aid in clustering in the case of K-means algorithm and should provide clear trends. Also based on the classes of each parameter, we can generate possible combinations and their effects, similar to a decision tree. A hybrid application framework[13] like Ionic will be used for the application using generated model.

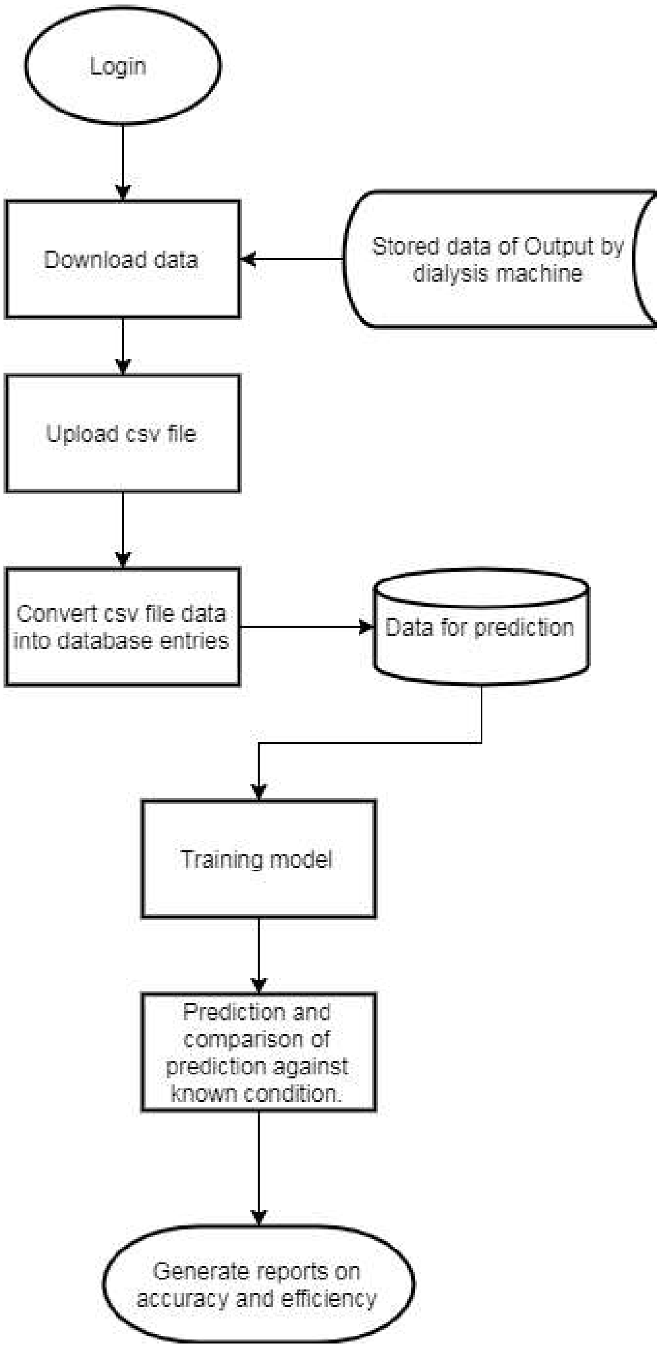


Fig. 3. Flow of system

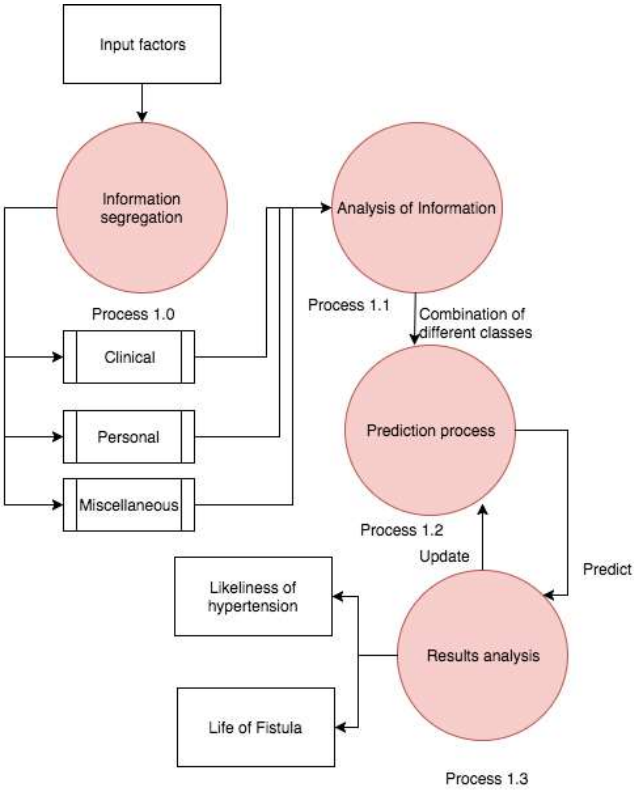


Fig. 4. Level 1 DFD

VII. Evaluation Details

All initial predictions made by the system must be compared with ultrasound doppler scans[5] to ensure that the system is generating accurate results.
Ai= 1 if result matches.
= 0 if no match
Accuracy = Summation(Ai)/ Count(A)
Where A is a matrix of comparison of output

VIII. Work done thus far

We have successfully identified important factors to be considered for the prediction system, modules that will be present in the system, flow between the modules and considered multiple algorithms for inclusion in the system. We have looked at the data of a failing fistula and a healthy one and used the same to understand the parameters better.
The results of the same are as follows:
The session 1 stats show an AVF which is failing. In Fig 5, the Session 1 blood flow is much under the recommended 300ml/min. Session 2 shows an acceptable blood flow, slightly above 300 ml/min.
The venous pressure in a healthy AVF should be increasing with duration of dialysis. In Fig 6, this is the case in Session 2, but not in Session 1 where it is dropping with time.
The UF speed should be as close to 1 as possible. In Fig 7, the low speed in Session 1 indicates that ultrafiltration is not

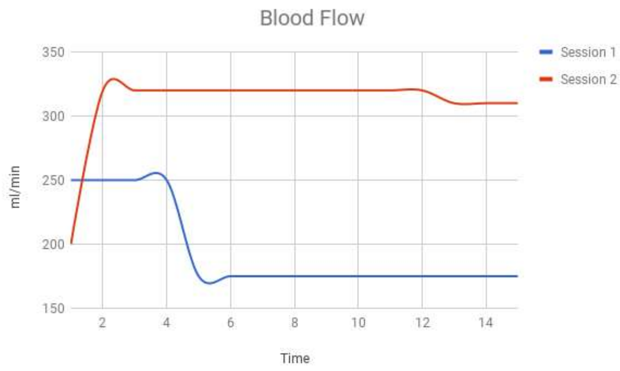


Fig. 5. Blood Flow

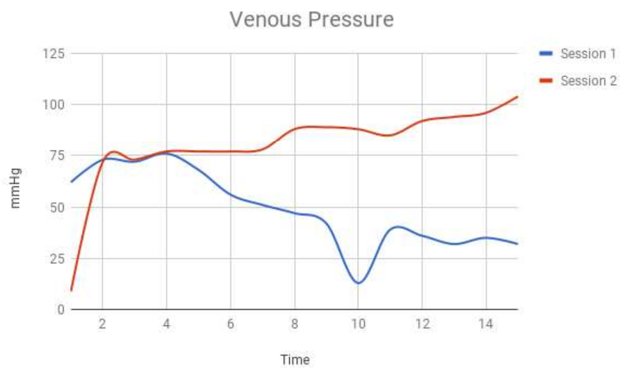


Fig. 6. Venous Pressure

Fig. 7. Ultrafiltration Speed

going at an acceptable rate.

The convection current in the artificial kidney creates a

Fig. 8. Trans Membrane Pressure

positive pressure difference allowing wastes to be filtered out. In Fig 8, this is not the case in Session 1. It is the case in Session 2.

The final kT/V value at the end of a dialysis session should

Fig. 9. Current kT/V

be at least 1.2, preferably higher. Fig 9 shows that in Session 1, the kT/V value is an ineffective 0.7. In Session 2 it is an effective 1.6.

IX. Conclusion

With the passage of time, the arteriovenous fistula is likely to fail. If or when this happens, the patient will be inconvenienced with having a catheter while a new fistula can be developed. This is extremely unfortunate because using the ultrasound doppler test [5] to check health of fistula is not feasible. Since dialysis is a continuous process, the dialysis machine is a treasure trove of information which can be used by our system to predict the health of the fistula without any expensive tests. Having this data declared at every dialysis session allows for continuous monitoring of the fistula over time. Thus if it is likely to fail, a new fistula can be created at another location and the current one can be used till it fails. In the meantime, the new fistula develops and dialysis can shift to it. This system when developed will improve quality of life of the patient due to foregoing the catheter while also reducing medical costs of scans like the ultrasound doppler[5].

References

[1] Homer William Smith, The kidney: structure and function in health and disease, Oxford University Press, USA Vol. 1, 1951.

[2] Jeremy Levy, Edwina Brown, Anastasia Lawrence, Oxford Handbook of Dialysis, Oxford University Press, Feb. 2016.

[3] Arif Asif, Prabir Roy-Chaudhury, Gerald A Beathard,Early arteriovenous fistula failure: a logical proposal for when and how to intervene, Clinical Journal of the American Society of Nephrology Vol. 1, issue 2, pages 332-339, March 2006.

[4] Rajiv Agarwal, Allen R Nissenson, Daniel Batlle, Daniel W Coyne, J Richard Trout, David G hemodialysis patients in the United States, The American journal of medicine Vol. 115, issue 4, pages 291-297, Sept. 2009.

[5] Pasquale Zamboli, Fulvio Fiorini, Alessandro DAmelio, Pasquale Fatuzzo, Antonio Granata, Color Doppler ultrasound and arteriovenous fistulas for hemodialysis, Springer International Publishing Vol. 17, Issue 4, pages 253-263, Dec. 2014.

[6] Norbert Lameire, Jean-Pierre Wauters, Jose Luis Grriz Teruel, Wim Van Biesen, Raymond Vanholder, An update on the referral pattern of patients with end-stage renal disease, Kidney International Vol. 61, pages S27-S34, May 2002.

[7] Manoj Reddy and John Cho, Detecting Chronic Kidney Disease Using Machine Learning, Qatar Foundation Annual Research Conference Proceedings Vol. 2016, issue 1, March, 2016.

[8] Enes Celik, Muhammet Atalay, Adil Kondiloglu, The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning, International Journal of Intelligent Systems and Applications in Engineering Vol. 4, issue special issue-1, pages 27-31, Dec. 2016.

[9] BV Ravindra, N Sriraam, M Geetha, Discovery of significant parameters in kidney dialysis data sets by K-means algorithm, IEEE, pages 452-454, Nov. 2014.

[10] Jeffrey Navarro Rojas, A Systematic Review on the Quality of Life among Hemodialysis Patients in the Middle East Countries, International Journal of Science and Research Vol. 6, issue 5, pages 532-542, May 2017.

[11] Martn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Man, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Vigas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv:1603.04467, March 2016.

[12] Congcong Li, Jie Wang, Lei Wang, Luanyun Hu, Peng Gong, Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery, Remote Sensing Vol. 6, issue 2, pages 964-983, Jan. 2014.

[13] Spyros Xanthopoulos, Stelios Xinogalos, A comparative analysis of cross-platform development approaches for mobile applications, ACM, pages 213-220, Sept. 2013.

11.3.2 Plagiarism report – Paper 1

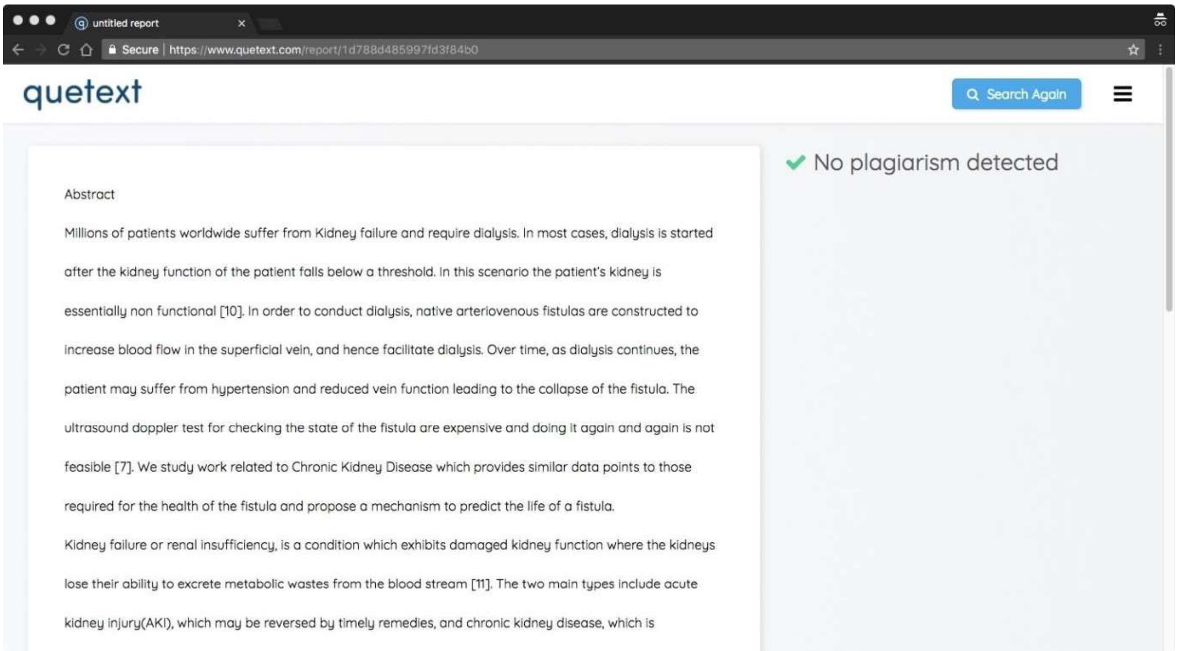


Fig 11.1 Section 1

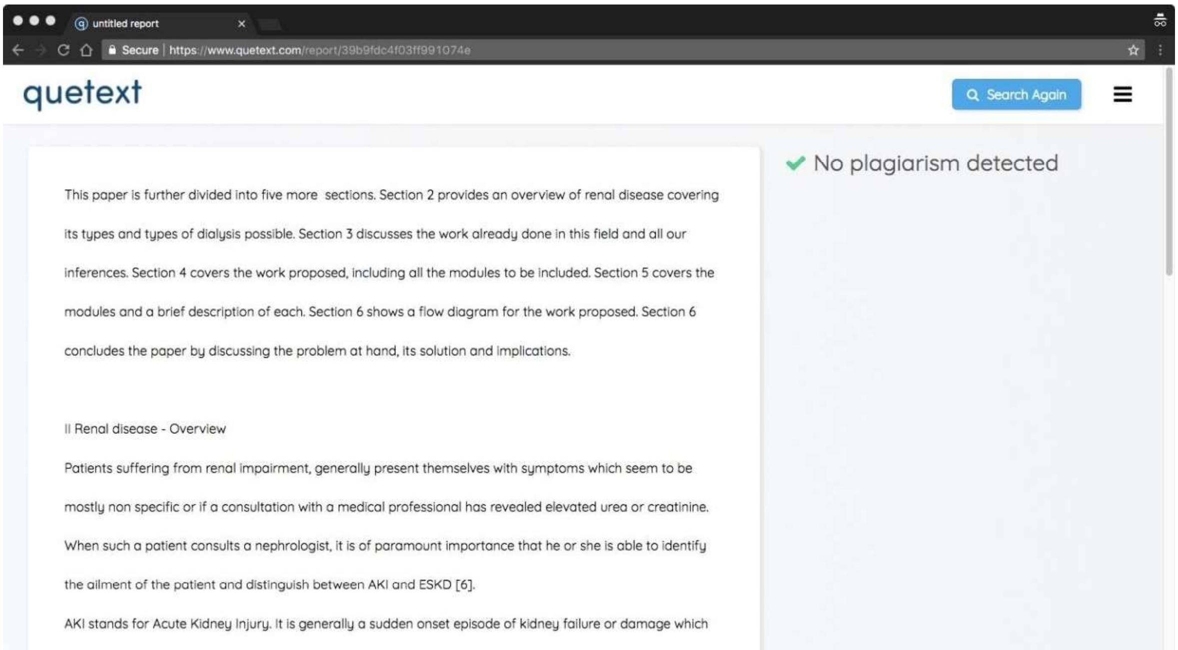


Fig 11.1 Section 2

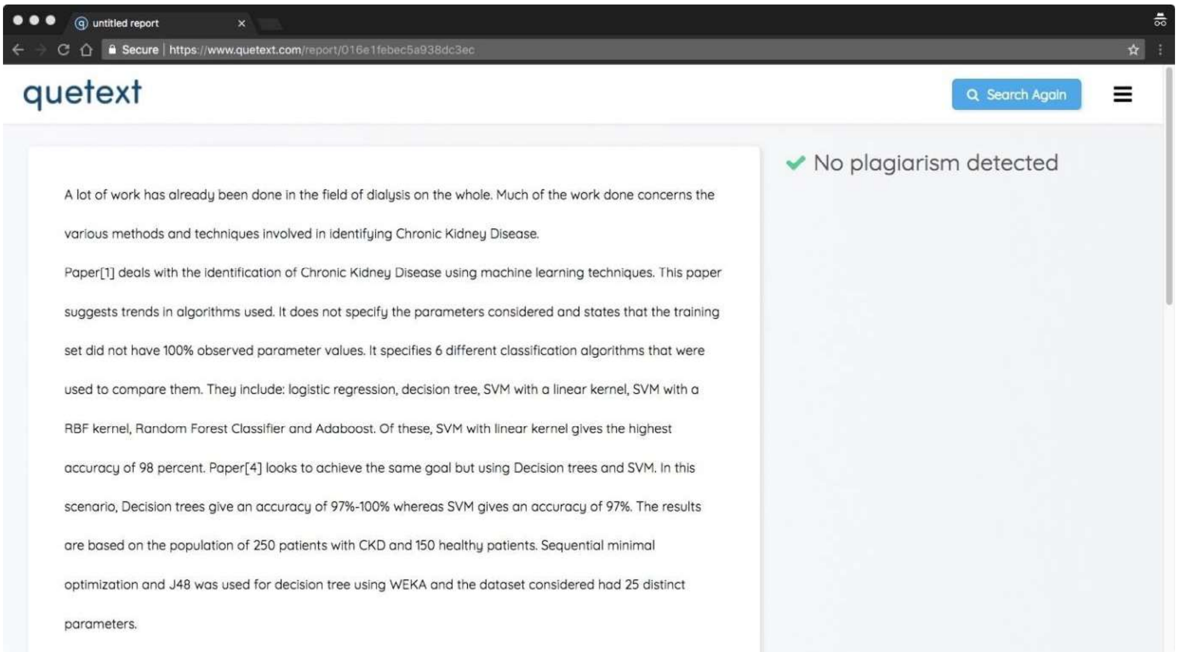


Fig 11.1 Section 3

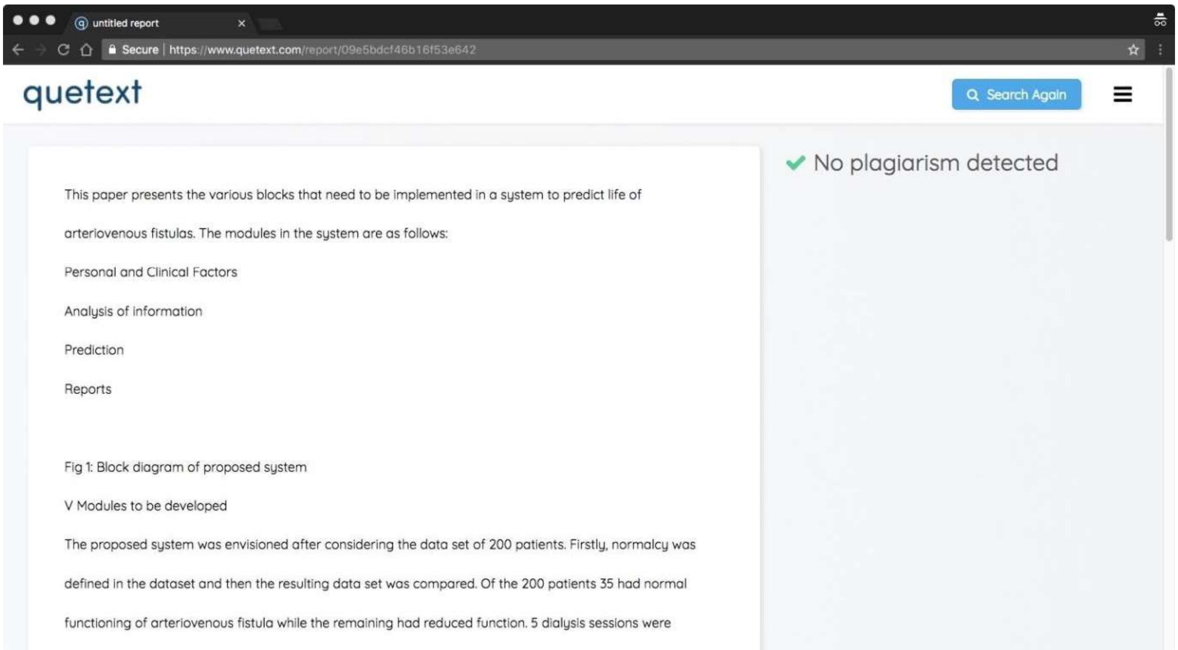


Fig 11.1 Section 4

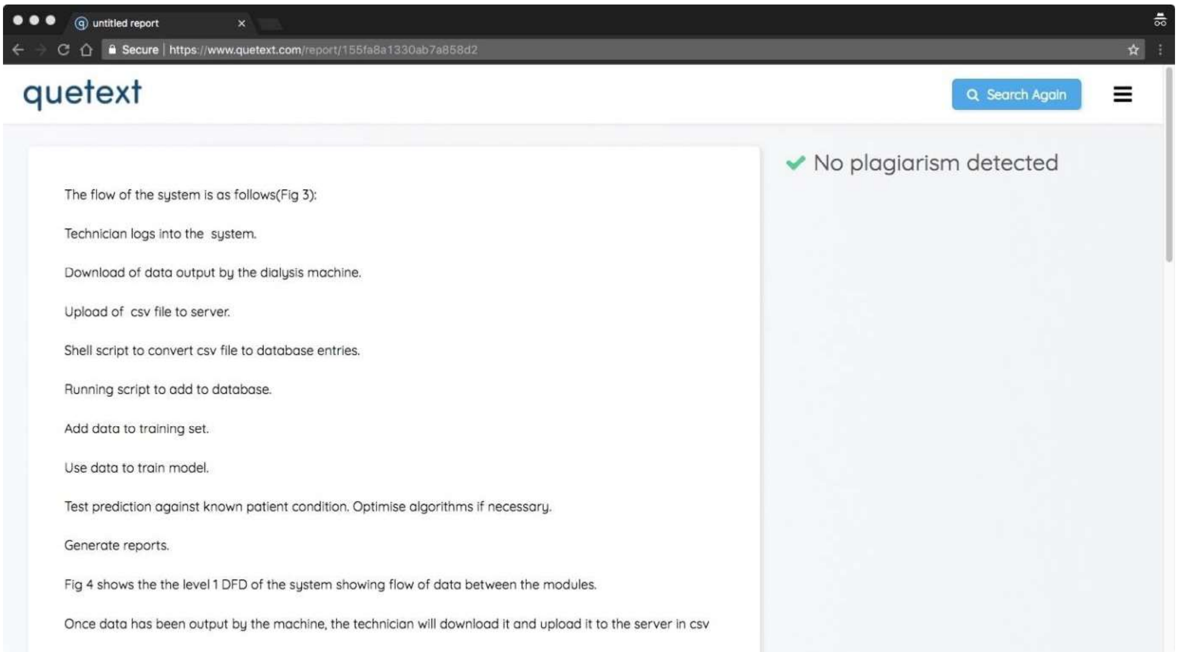


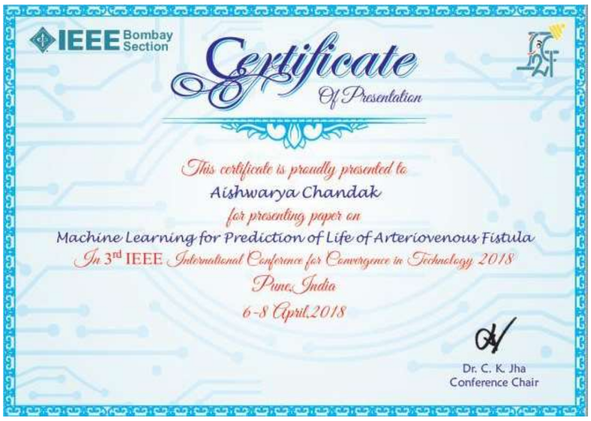
Fig 11.1 Section 5

11.3.3 Certificates– Paper 1


IEEE Bombay Section
Certificate
Of Presentation
This certificate is proudly presented to
Mihir Wagle
for presenting paper on
Machine Learning for Prediction of Life of Arteriovenous Fistula
In 3rd IEEE International Conference for Convergence in Technology 2018
Pune, India
6-8 April,2018
Dr. C. K. Jha
Conference Chair


IEEE Bombay Section
Certificate
Of Presentation
This certificate is proudly presented to
Neeraj Jethnani
for presenting paper on
Machine Learning for Prediction of Life of Arteriovenous Fistula
In 3rd IEEE International Conference for Convergence in Technology 2018
Pune, India
6-8 April,2018
Dr. C. K. Jha
Conference Chair


IEEE Bombay Section
Certificate
Of Presentation
This certificate is proudly presented to
Juhi Bhagtani
for presenting paper on
Machine Learning for Prediction of Life of Arteriovenous Fistula
In 3rd IEEE International Conference for Convergence in Technology 2018
Pune, India
6-8 April,2018
Dr. C. K. Jha
Conference Chair


IEEE Bombay Section
Certificate
Of Presentation
This certificate is proudly presented to
Aishwarya Chandak
for presenting paper on
Machine Learning for Prediction of Life of Arteriovenous Fistula
In 3rd IEEE International Conference for Convergence in Technology 2018
Pune, India
6-8 April,2018
Dr. C. K. Jha
Conference Chair


IEEE Bombay Section
Certificate
Of Presentation
This certificate is proudly presented to
Gresha Bhatia
for presenting paper on
Machine Learning for Prediction of Life of Arteriovenous Fistula
In 3rd IEEE International Conference for Convergence in Technology 2018
Pune, India
6-8 April,2018
Dr. C. K. Jha
Conference Chair

11.3.4 Paper 2 – presented at DJ ASCII 2018

Application of ML techniques for the analysis of hypertension and prediction of vein function in hemodialysis

Dr. Gresha Bhatia	Mihir Wagle	Neeraj Jethnani	Juhi Bhagtani	Aishwarya Chandak
Deputy HOD, CMPN Dept.	Student, CMPN Dept.	Student, CMPN Dept.	Student, CMPN Dept.	Student, CMPN Dept.
V.E.S.I.T	V.E.S.I.T	V.E.S.I.T	V.E.S.I.T	V.E.S.I.T
Chembur, Mumbai	Chembur, Mumbai	Chembur, Mumbai	Chembur, Mumbai	Chembur, Mumbai

Domain

Artificial Intelligence-Machine Learning

Abstract

Millions of patients worldwide suffer from Kidney failure and require dialysis. In most cases, dialysis is started after the kidney function of the patient falls below a threshold. In this scenario the patient's kidney is essentially non functional. In order to conduct dialysis, native arteriovenous fistulas are constructed to increase blood flow in the superficial vein, and hence facilitate dialysis. Over time, as dialysis continues, the patient may suffer from hypertension and reduced vein function leading to the collapse of the fistula. The ultrasound doppler test for checking the state of the fistula is expensive and doing it again and again is not feasible. The proposed work explores the Chronic Kidney Diseases and proposes a mechanism that uses optimised data points to predict health of fistula.

Problem Definition

Predict the state of the native arteriovenous fistula considering several attributes like venous pressure, arterial pressure, systolic blood pressure, diastolic blood pressure, heart rate, etc

Literature Survey

A lot of research has been done in the field of dialysis. The focus lies in various methods and techniques involved in identifying Chronic Kidney Disease.

Reddy and co authors, in their paper deal with the identification of Chronic Kidney Disease using Machine learning techniques. This paper suggests trends in algorithms used. It does not specify the parameters considered. It specifies 6 different classification algorithms that were used to compare them. Of the algorithms used, SVM with linear kernel gives the highest accuracy of 98 percent.

Celik and co authors, in their paper look to achieve the same goal but only using Decision trees and SVM. In this scenario, Decision trees and SVM give an accuracy of 97%-100% and 100% respectively. The results are based on the population of 250 patients with CKD and 150 healthy patients. Sequential minimal optimization and J48 was used for decision tree using WEKA and the dataset considered had 25 distinct parameters.

Ravindra and co authors, in their paper try to find significant parameters in kidney dialysis sets using the K-means algorithm. It relies on classifying parameters into ranges such as medium, low and high to further aid with clustering. This paper mainly focused on identifying survival period of patient undergoing dialysis using clustering techniques. In this scenario, creatinine plays an important part and it is found that patients with a level of creatinine which is either high or low suffer from adverse effects.

Rojas and co authors, deal with the effects of dialysis to the quality of life of a patient. It shows the results of a survey of patients undergoing hemodialysis. Starting hemodialysis involves a significant lifestyle change and can have a lot of effects on a patient's physical and mental health. Care needs to be taken that changes are not averse. The paper was restricted to specific region and results could not be generalized.

Based on the manual of the Nipro Surdial 55 Plus, a common dialysis machine, it is observed that it outputs a variety of factors on its display, updating second on second. Most of this data is considered in the present and not stored. However, the same can be done since the machines allow backup to a database.

Looking at all the related work it was identified that identification and classification is an important aspect of work in the field of dialysis. However, there is a need to provide for forecasting and prediction of life of arteriovenous fistula by making use of state of the art techniques such as Machine Learning and Big data.

The research papers surveyed and interactive sessions with nephrologists have furthered our understanding of the dialysis process and its subprocesses and motivated us to come up with relevant factors to be considered in relation to health of the fistula and work on the proposed problem statement.

Proposed Architecture with modular description

The various blocks that are implemented in the proposed system to predict life of arteriovenous fistulas are shown in the figure 1. The modules in the system are as follows:

a. Personal and Clinical Factors
b. Analysis of information
c. Prediction
d. Reports

The proposed system was envisioned after considering the data set of 200 patients. Firstly, normalcy was defined in the dataset and then the resulting data set was compared. Of the 200 patients 35 had normal functioning of arteriovenous fistula while the remaining had reduced function. 5 dialysis sessions were included for each patient to avoid any case of operator error.

The dataset comprises of over 30 values reported by the machine. Some of them are:

a. Dialysis flow
b. Blood flow
c. Arterial Pressure
d. Venous Pressure
e. Ultrafiltration speed
f. Heart rate
g. Trans membrane pressure
h. Total Ultrafiltration speed
i. Blood pressure
j. Target kT/V

There are four distinct modules need to be developed in the system in order to effectively predict the life of arteriovenous fistula:

Personal & Clinical data: This module takes in data which is being output from the dialysis

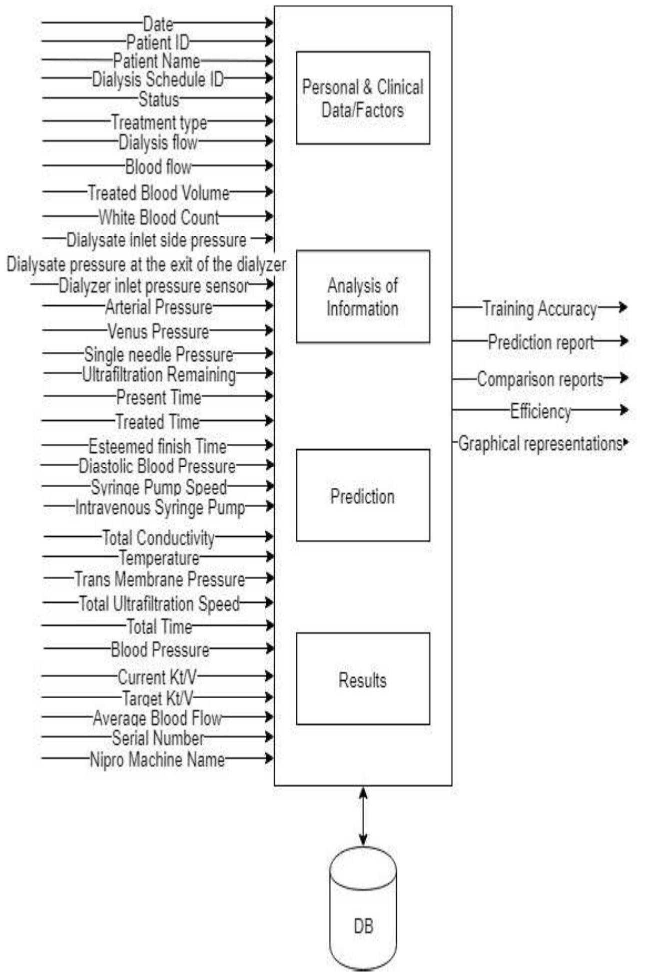


Fig 1: Architecture for the proposed system

machine and stores it within a database after segregating it based on the type of data. Data may be divided into the categories - personal, clinical or miscellaneous.

Analysis of Information: This module will take in data from the database, create graphs based on the same and analyze them to see which factors are significant in predicting health of fistula

Prediction: This module will consider the following algorithms: SVM, RFC, MLP. We will check their efficacy on the data at hand. Based on the efficiency new algorithms may be developed to improve accuracy of the system.

Results: In this module, comparison of the predicted value of the state of fistula with the actual condition is performed to calculate the accuracy of the system and generate reports to track performance of the system.

Expected Outcome

The expected outcome comprises of a report consisting of all the factors considered for predicting the state of arteriovenous fistula, state of the fistula, comparative study of the accuracy of the algorithms used, graphical representations of accuracy, prediction, false positives and false negatives. The patient will be sent for an ultrasound doppler test only if the report states his/her fistula as an unhealthy fistula .

11.3.5 Plagiarism report – Paper 2

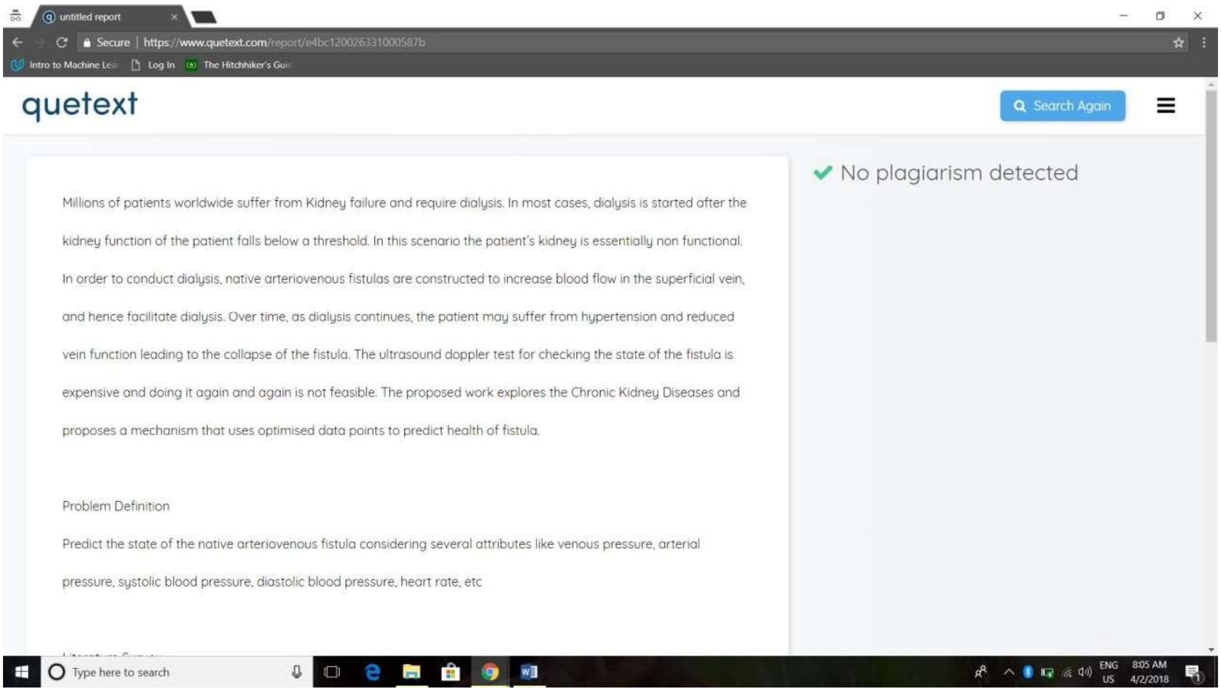


Fig 9.2 Section 1

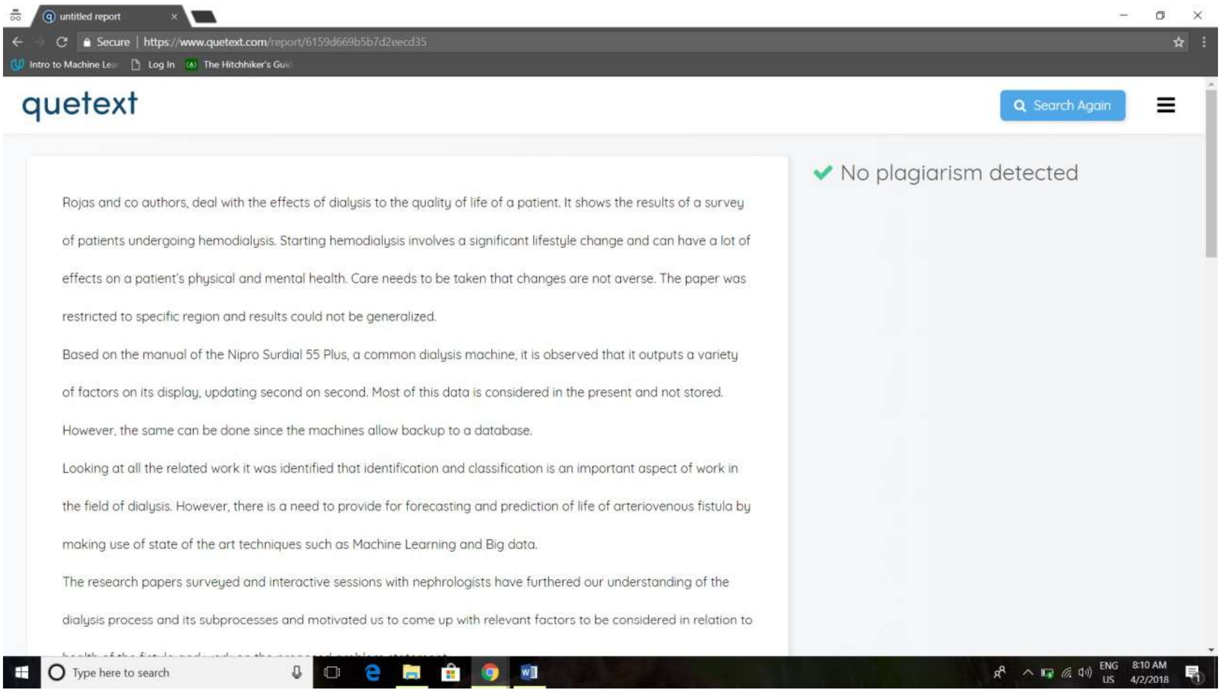


Fig 9.2 Section 2

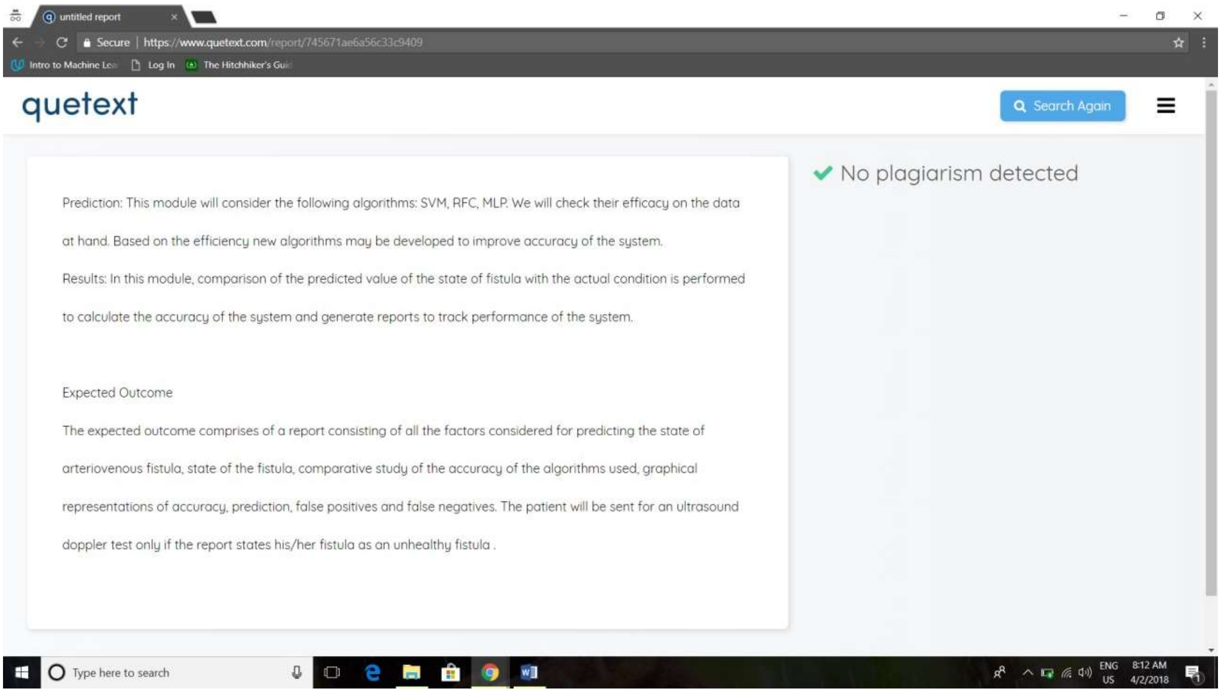


Fig 9.2 Section 3

11.3.6 Certificate– Paper 2

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
Vile Parle (W) Mumbai - 400 056.

DJASCII
(State level Project Competition)

Certificate of Participation

This is to certify that MIHIR WAGLE has participated in the State level Project Competition - DJASCII, conducted by Dwarkadas J. Sanghvi College of Engineering, Mumbai on 12th April, 2018.

Dr. N. M. Shekokar	Dr. Manali J. Godse	Dr. A. C. Daptardar	Dr. Hari Vasudevan
Joint-Convener Prof. & Head, Dept. of Comp. Engg.	Vice Principal (Acads.) Prof. & Head, Dept. of Bio-Medical Engg.	Vice Principal (Admin.)	Convener Principal

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
Vile Parle (W) Mumbai - 400 056.

DJASCII
(State level Project Competition)

Certificate of Participation

This is to certify that NEERAJ JETHNANI has participated in the State level Project Competition - DJASCII, conducted by Dwarkadas J. Sanghvi College of Engineering, Mumbai on 12th April, 2018.

Dr. N. M. Shekokar	Dr. Manali J. Godse	Dr. A. C. Daptardar	Dr. Hari Vasudevan
Joint-Convener Prof. & Head, Dept. of Comp. Engg.	Vice Principal (Acads.) Prof. & Head, Dept. of Bio-Medical Engg.	Vice Principal (Admin.)	Convener Principal

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
Vile Parle (W) Mumbai - 400 056.

DJASCII

(State level Project Competition)

Certificate of Participation

This is to certify that JUHI BHAGTANI has participated in the State level Project Competition - DJASCII, conducted by Dwarkadas J. Sanghvi College of Engineering, Mumbai on 12th April, 2018.

Dr. N. M. Shekokar	**Dr. Manali J. Godse**	**Dr. A. C. Daptardar**	**Dr. Hari Vasudevan**
Joint-Convener	Vice Principal (Acads.)	Vice Principal (Admin.)	Convener
Prof. & Head, Dept. of Comp. Engg.	Prof. & Head, Dept. of Bio-Medical Engg.		Principal

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
Vile Parle (W) Mumbai - 400 056.

DJASCII

(State level Project Competition)

Certificate of Participation

This is to certify that AISHWARYA CHANDAK has participated in the State level Project Competition - DJASCII, conducted by Dwarkadas J. Sanghvi College of Engineering, Mumbai on 12th April, 2018.

Dr. N. M. Shekokar	**Dr. Manali J. Godse**	**Dr. A. C. Daptardar**	**Dr. Hari Vasudevan**
Joint-Convener	Vice Principal (Acads.)	Vice Principal (Admin.)	Convener
Prof. & Head, Dept. of Comp. Engg.	Prof. & Head, Dept. of Bio-Medical Engg.		Principal