**DETERMINATION OF GENUINENESS OF A MEDICAL PROFESSIONAL**

**Blank Sheets:**

In addition to the white sheets (binding requirement) two white sheets shall be put at the beginning and the end of the thesis.

**DETERMINATION OF GENUINENESS OF A MEDICAL PROFESSIONAL**

Submitted in partial fulfillment of the requirements

of the degree of

**B. E. Computer Engineering**

By

**Aayush Shah 60004140091**

**Pranay Shah 60004140101**

**Dipam Vasani 60004140116**

Guide(s):

**Kriti Srivastava**

Assistant Professor

|  |  |  |
| --- | --- | --- |
| D:\Admission\Admissions_1011\wwwroot\hdrlogo.gif | Department of Computer Engineering  D. J. Sanghvi College of Engineering  Mumbai – 400 056 | E:\new logo.JPG |

University of Mumbai

2017-2018

**CERTIFICATE**

This is to certify that the project entitled **“Determination of genuineness of a medical professional”** is a bonafide work of **“Aayush Shah” (60004140091), “Pranay Shah” (60004140101), “Dipam Vasani” (60004140116)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of B.E. in Computer Engineering

**Prof. Kriti Srivastava**

**Internal Guide**

**Dr. N. M. Shekokar Dr. Hari Vasudevan**

**Head of Department Principal**

i

**Project Report Approval for B.E.**

This project report entitled ***Determination of genuineness of a medical professional*** by ***Aayush Shah, Pranay Shah, Dipam Vasani*** is approved for the degree of ***B.E. in Computer Engineering.***

Examiners

1.---------------------------------------------

2.---------------------------------------------

Date:

Place:

ii

Declaration

I/We declare that this written submission represents my/our ideas in my/our own words and where others' ideas or words have been included, I/We have adequately cited and referenced the original sources. I/We also declare that I/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 my/our submission. I/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)

-----------------------------------------

(Name of student and Roll No.)

Date:

iii

**ABSTRACT**

In the age, where any data is easily available, it is of utmost importance that the data accessed by a person, isn’t used for malicious purposes, or more importantly the data isn’t in the wrong hands. This is even more important when the data is about an individual’s health and medical history. Often, we hear about cases wherein a medical professional misused a person’s past medical history. So, we thought of determining a medical professional’s genuineness, and then designing a system that may then decide to deny access to a malicious doctor. These results would solely be based upon the credulity of a medical professional. In this work, we discuss the design and implementation of the proposed project, using supervised Neural Networks. A knowledge base would be developed for this purpose to help determine whether a doctor is genuine or malicious. Factors like the location from which a medical professional accesses data, relevance of the data being accessed to the ailment/treatment, etc. would serve as inputs to the Neural Network, and using Error Back Propagation Technique, the Network would be trained for a variety of inputs Finally, the genuineness of a medical professional as a numerical value between zero and one, and access would be granted or denied based on a predetermined threshold.

iv

**Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Contents** | **Page No.** |
| **1** | **INTRODUCTION** | **1** |
|  | **1.1 Description** | **2** |
|  | **1.2 Problem Formulation** | **2** |
|  | **1.3 Motivation** | **3** |
|  | **1.3 Proposed Solution** | **4** |
|  | **1.4 Scope of the project** | **6** |
| **2** | **REVIEW OF LITERATURE** | **7** |
| **3** | **SYSTEM ANALYSIS** | **9** |
|  | **3.1 Functional Requirements** | **9** |
|  | **3.2 Specific Requirements** | **10** |
|  | **3.3 Use-Case Diagrams and description** | **11** |
| **4** | **ANALYSIS MODELING** | **13** |
|  | **4.1 Activity Diagram** | **13** |
|  | **4.2 Functional Modeling** | **14** |
|  | **4.3 TimeLine Chart** | **15** |
| **5** | **DESIGN** | **16** |
|  | **5.1 Architectural Design** | **16** |
| **6** | **CONCLUSION AND FUTURE SCOPE** | **17** |
| **7** | **ACKNOWLEDGEMENT** | **18** |

v

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Caption** | **Page No.** |
| 3.3.1 | Use case diagram | 11 |
| 4.1.1 | Activity diagram | 13 |
| 4.2.1 | Data flow diagram | 14 |
| 4.3.1 | Timeline diagram | 15 |
| 5.1.1 | System Architecture diagram | 16 |

vi

**Chapter 1**

**INTRODUCTION**

Medical records kept in a hospital can be used as a personal or impersonal document. Personal document involves information that is confidential and should not be released without the consent of the patient except in some specific situations. Impersonal document are those record with a loose identity as a personal document and patient permission is not required. These records could be used for research purposes. Confidentiality is an important component of the rights of the patient. The hospital is legally bound to maintain the confidentiality of the personal medical records. The patient can claim negligence against the hospital or the doctor for a breach of confidentiality. However, there are certain situations where it is legal for the authorities to give patient information. These include during referral, when demanded by the court or by the police on a written requisition, when demanded by insurance companies as provided by the Insurance Act when the patient has relinquished his rights on taking the insurance, and when required for specific provisions of Workmen's Compensation cases, Consumer Protection cases, or for Income tax authorities. The maintenance of confidentiality is an important issue in the era of electronic data storage. There should be checks in place so that only those who are authorized can access the patient data.

However, an authorized individual may misuse the data too for malicious purposes. There have been incidents where doctors have accepted bribes to leak sensitive patient information such as HIV status, mental health history and abortions to unauthorized personnel. There have also been incidents where mediclaim information is leaked by doctors to insurance companies for a commission, and later the insurance agencies use this information for unlawful purposes. In order to prevent such scenarios we have come up with a system to find out the genuineness of a medical professional. Data access can be denied by the system to a doctor who is found to be malicious.

**1.1 DESCRIPTION**

This main purpose of this project to find the credulity of a medical professional for a hospital management system. When a doctor accesses a patient’s data, we determine whether or not the doctor may misuse this information for any malicious purpose. We can do this by looking at various cases such as the time and location of the data access. The access is likely to be genuine if the patient’s data is accessed within hospital hours and inside the hospital. Now, we can also refer to the type of data access and then map it to his profession. Ideally, a doctor will not access information that is not relevant to the area of his specialization. If he does so then there is a chance that he is malicious. We have used all these factors to determine whether the doctor is malicious or genuine. Data access can now be denied to a doctor who is found to be malicious.

**1.2 PROBLEM FORMULATION**

Malicious doctors in hospitals often indulge in leaking a patient’s data to people outside the hospital. A few of these are mentioned subsequently. Medical researchers, including pharmaceutical companies, are regularly given medical records, including names. Medical students often get the name and medical history - and even tissue samples - of living patients, without their consent or knowledge. A 68-year-old man was refused a place in a care home when social services found from his medical records that he was gay. An uncle found out that his niece had a secret abortion when the company he worked for was asked to do a financial audit of the local health authority. He told her parents, who are very religious. A woman was sacked after her GP sent her records to her employer. The notes revealed that she had a history of mental health problems. Patients with medical conditions have been approached by researchers who have had access to their records. An MP was sent the medical records of a constituent without her consent. She found out only when the MP passed on the records to her. Our goal is to prevent such situations from happening. This can be done by finding the genuineness of a doctor and then blocking access to him if he is found to be malicious.

**1.3 MOTIVATION**

There are several methods to determine the credulity of a medical professional. In traditional access control systems, security administrators determined whether an information consumer can access a certain resource. However, in reality, it is very difficult for policy makers to foresee what information a user may need in various situations. In hospitals, failing to authorize a doctor for the medical information he needs about a patient could lead to severe or fatal consequences.

An existing method used in hospitals involves categorizing medical records based on content, and then evaluating if a doctor is malicious based on what content he accesses. However, the drawback in this approach is that a doctor’s specialization affects the type of data he accesses. For example, doctors specializing in sexually transmitted diseases (STD) will naturally need a lot of STD related information in their work. STD records are preserved to be more sensitive than many other medical records. If risk scores are determined by data sensitivity, STD specialists would aggregate risk much faster than other doctors.

A quick method to overcome our problem would be to apply a clustering algorithm based on the data accessed and then apply outlier detection algorithms to find malicious doctors. However, the principal assumption with this method is that the number of doctors who are genuine should be far greater than those who are malicious. If the opposite was to be true, then a cluster of malicious doctors would be formed and a genuine doctor would be found to be an anomaly. This should not be the case.

To overcome these drawbacks, we have come up with an innovative solution that involves using a neural network to find out if a doctor is malicious or not. We have used a rule base to map a doctor’s profession to the type of data he requires and we have used this as a basis to determine whether the data he has accessed is used for genuine or malicious purposes.

**1.4 PROPOSED SOLUTION**

We have incorporated a neural network in our project that determines if a doctor is malicious or genuine by providing an output between zero and one. Higher the value of the final output of our system, the greater the chances of him being malicious. We have used three main inputs to our network which are time of data access, the location of access, and the relevancy of the data accessed to the doctor’s profession.

We have defined a rule base that helps us decide the weights of each input. If a doctor accesses a patient’s data outside hospital hours, then there is a chance that he is malicious. Alternatively, if data is accessed within hospital hours there are no signs to show he is malicious. Moving on, if a doctor accesses data from his personal computer in the hospital, or from another machine in the hospital then there are less chances of him being malicious. However, if the data is accessed from a machine outside the hospital then it can be used for malicious purposes. Another important factor to consider is the relevance of different types of data to a doctor’s profession. We have conducted extensive research to find out the data needed for five major specializations and incorporated this information into our project. We have also incorporated provisions for an emergency scenario in our project, where a doctor may access information outside the hospital that is not relevant to his specialization but needs urgent diagnosis. In such a scenario access should not be denied to a doctor.

We are looking to train our neural network with several examples of genuine and malicious doctors based on the rule base that we have created. We also plan to use this rue base to define the initial weights of each and every neuron in our model and use appropriate activation functions to calculate the final output of our network. We are planning to use backpropogation technique to correct the errors in initial weights and help the machine learn. Finally, we will have a neural network that will be able to efficiently predict whether a doctor requesting data is genuine or malicious. Thus by obtaining this information, the system can appropriately decide whether it should grant patient information to a particular doctor, or if we should deny this information to a doctor.

Our innovative approach eliminates the drawbacks mentioned in earlier approaches. We have taken into consideration various factors that vary among doctors and may change from time to time. We have efficiently determined how much risk we should tolerate for a doctor and have correspondingly designed our access control system.

Our system will not produce any erroneous output that may have been the case for a system that used a clustering approach followed by an outlier detection algorithm for anomaly detection. Since the neural network will be trained with a large number of examples that use a well detailed rule base as a basis for the learning processes, we will have a neural network that will be very accurate. After several iterations of learning by weight adjustment, it will be able to produce an accurate result that is free from errors.

To summarize, our solution efficiently enforces learning techniques that are based on relevance of information to a doctor’s profession and take into account the time and location of data access. By taking an emergency case into consideration too we have a precise system that can correctly predict if a doctor is genuine or malicious. Based on the access can be denied to a malicious doctor.

**1.5 SCOPE**

After extensive research that involved meeting professional medical practitioners and looking at medical research publications we have designed our rule base taking into consideration five major specializations. These include gynecologists, dentists, psychiatrists, dermatologists and ophthalmologists. Our system is designed to categorize malicious or genuine doctors that fall under these specializations.

Due to a patient’s medical records not being readily available online, we are designing our own data set for this project. This will be done by referring a few sample data sets online and then forming our own data set. Thus, the neural network will be trained as per the examples from the data set develop by us. We have to go for this approach since data is not available freely in our desired form as no hospitals keep their data open source. If we get access to a data set in the form needed by us, we can use that to train our system.

A future scope of this project could be to increase the areas of specializations and perform additional research to accordingly update our rule base. The neural networks will have be trained again with examples from these additional specializations for it to find out the credulity of a doctor belonging to that profession.

**Chapter 2**

**REVIEW OF LITERATURE**

There are not many methods proposed by researchers that help restrict or in that case prevent any fraud against the misuse of a patient’s private data. While a lot of research has been done on protecting an IT system against unauthorized access from attackers, not much work has been done on preventing sensitive information from being jeopardized, either due to abuse or carelessness, by authorized users. And since it is extremely critical that we protect sensitive data against any kind of malicious intent, there is an increasing need for a more advanced, and highly secure and accurate system.

There was a proposed system [1] which used a quantified-risk adaptive approach in order to keep a check on the access of a patient’s privacy. The paper claims that even though doctors are authorized to access their patient’s data, in reality, there is also an inherent risk in each access. Their proposed solution therefore, allows information consumers to have the freedom to choose what they want to access. A user’s data-accessing activities are associated with quantified risk scores, which will be added up over time. Request to access a resource is granted if doing so will not make the user’s aggregated risk exceed his/her tolerance threshold set by the system; otherwise, the request is denied. They do this by calculating a relevance-relation function. They determine what activity a certain honest doctor will perform with respect to a disease, hence calculating the probability that a certain record m will be accessed to serve purpose p. They also take special cases into consideration. Once the model is trained they expect that a malicious doctor would over access a patient’s data hence increasing the risk considerably.

The problem with this method is that it is not as accurate as one expects it to be. Another glaring flaw is that they have only considered one parameter to determine if a doctor is malicious, i.e. relevance of the data accessed. The solution thus ends up ignoring myraid of other factors that can contribute towards determining spiteful activities, thus proving to be inefficient.

In another paper [2], clustering algorithms were used to analyse and hence group hospital data for better management. The paper mainly focuses on order history. It processes these order histories in order to find out the temporal global characteristics of clinical activities. Once it has done so, it keeps applying clustering techniques to the results until they converge. The final output is then expected to be the optimum output. For example the output may be a particular day when there is highest activity for a particular test or the time of the day when there is least activity. The major drawback of a clustering based approach is that, since it is unsupervised, the clusters formed may be the opposite of ones we expect. For example if there were a lot of malicious overprescribed orders in the history database, those will form a cluster and hence end up as a category.

The comparison table of the studied reference papers is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Reference Paper | Model | Drawbacks |
| [1] | Quantified Risk-Adaptive Access Control for Patient  Privacy Protection in Health Information Systems | A risk based adaptive model to keep a check on the data consumed by information consumers. | The system is less accurate and considers only one parameter i.e. relevance of data accessed. |
| [2] | Clustering-based Analysis in Hospital Information Systems | A clustering based algorithm to group orders based on various identified attributes | Clustering may lead to a lot of fake orders forming a cluster. |

**Chapter 3**

**SYSTEM ANALYSIS**

**3.1 Functional Requirements**

* **Capturing Details about the access**
* **Calculate the initial weights**
* **Refer to the Doctor’s Previous History**
* **Grant or Deny Access**
  + 1. **Capturing Details about the access**

Details about the access are captured in this step. These details include factors like time of the access, location of the access in terms of the device used to access the data, relevancy of the data being accessed is recorded, this relevance is based on if the data being accessed is relevant to the patient’s illness or the doctor’s specialization, also, if the data being accessed is in emergency circumstances or not, is also tested, as during such circumstances too much data might be required at once.

* + 1. **Calculate the initial weights**

Based upon the above details captured, initial weights for the neural network are calculated. These weights are crucial for the accuracy of the results that are obtained in the due course of the project.

* + 1. **Refer to the Doctor’s Previous History**

Referring to a Doctor’s precious history includes, checking only the data access grant/deny history of a doctor. If a Doctor was denied data access in the past, then the probability of him being denied access in the future is significant too.

* + 1. **Grant or Deny Access**

Finally, based on the above parameters, a decision is taken, whether to grant or deny the requested data’s access to the Doctor.

* 1. **Specific Requirements**

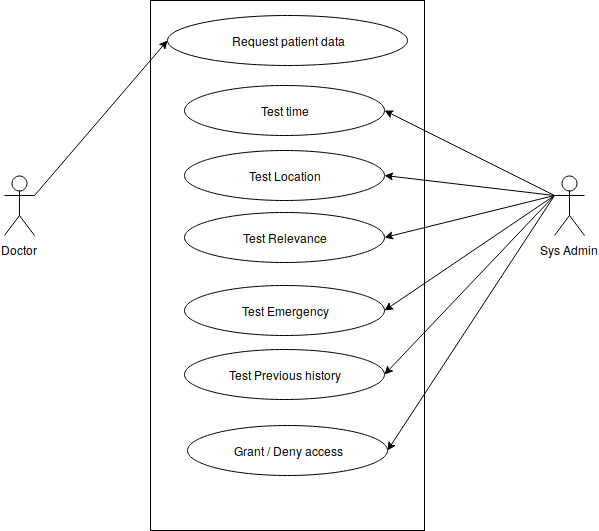
Our project would work best with the following hardware and software requirements, and these, would differ for a System Administrator and a Doctor.

For a System Administrator the corresponding requirements would be:

* A high-performance computer with at least 4GB RAM.
* MATLAB software
* ANFIS Editor
* High-speed internet connection
* A large database to store all the relevant information, and provide it as and when required

For a Doctor, the only requirement is a good, high-speed internet connection, that can send access request quickly to the System Administrator, and receive the corresponding result at the earliest.

**3.3 Use-Case Diagram and Description**



**Fig 3.3.1 Use- Case Diagram**

**Use- Case:** The users here are of 2 types:

* The Doctor, and
* The System Administrator

Each of them, perform different tasks.

**Description:**

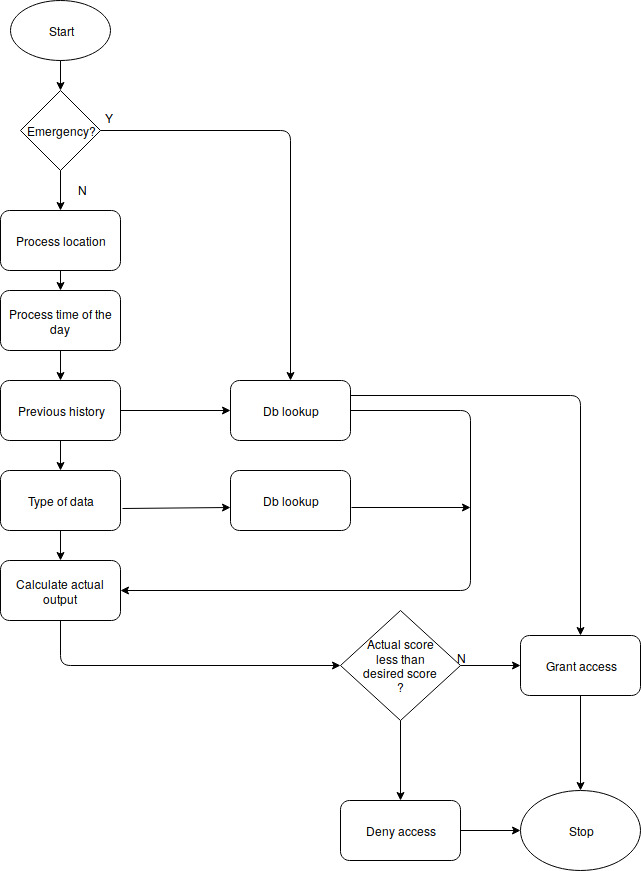
The Doctor can request for a patient’s data, as per his/her requirements.

The System Administrator, based upon factors like the time of access, location of access, relevance of the data requested, previous history, etc. would decide if to grant or deny access data to the doctor requesting it. If it is an emergency case, this factor precedes other factors in terms of priority or weightage and only the doctor’s previous history data is accessed to decide if access to the required data should be granted or not.

**Chapter 4**

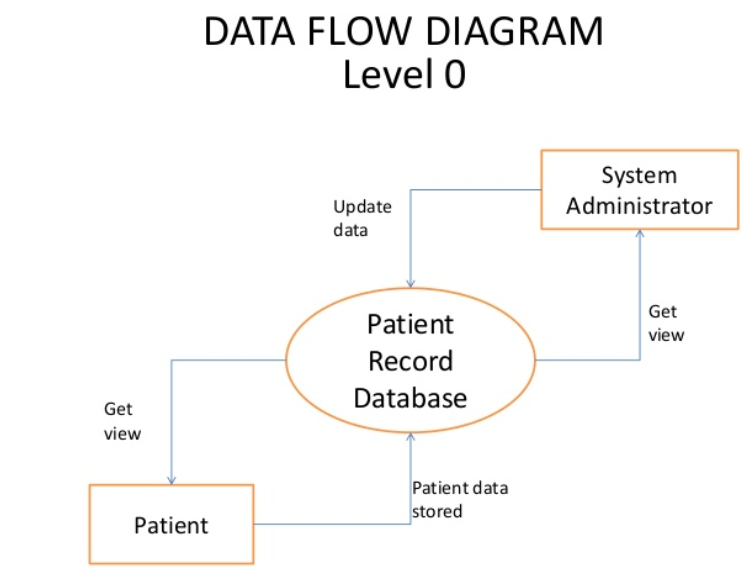
**ANALYSIS MODELING**

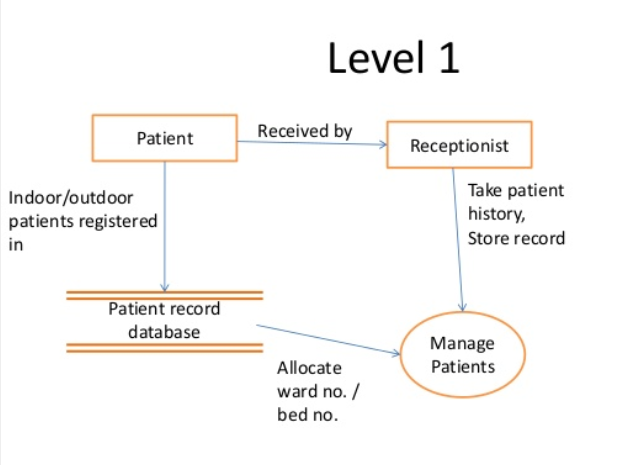
**4.1 Activity Diagram**

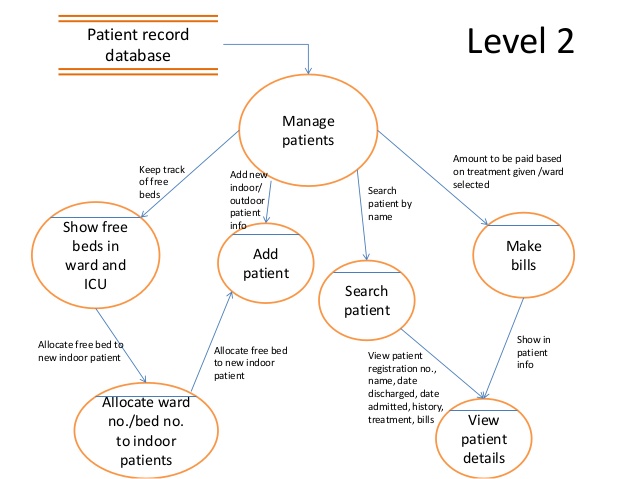
****

**Fig 4.1.1 Activity Diagram**

**4.2 Function Modeling**

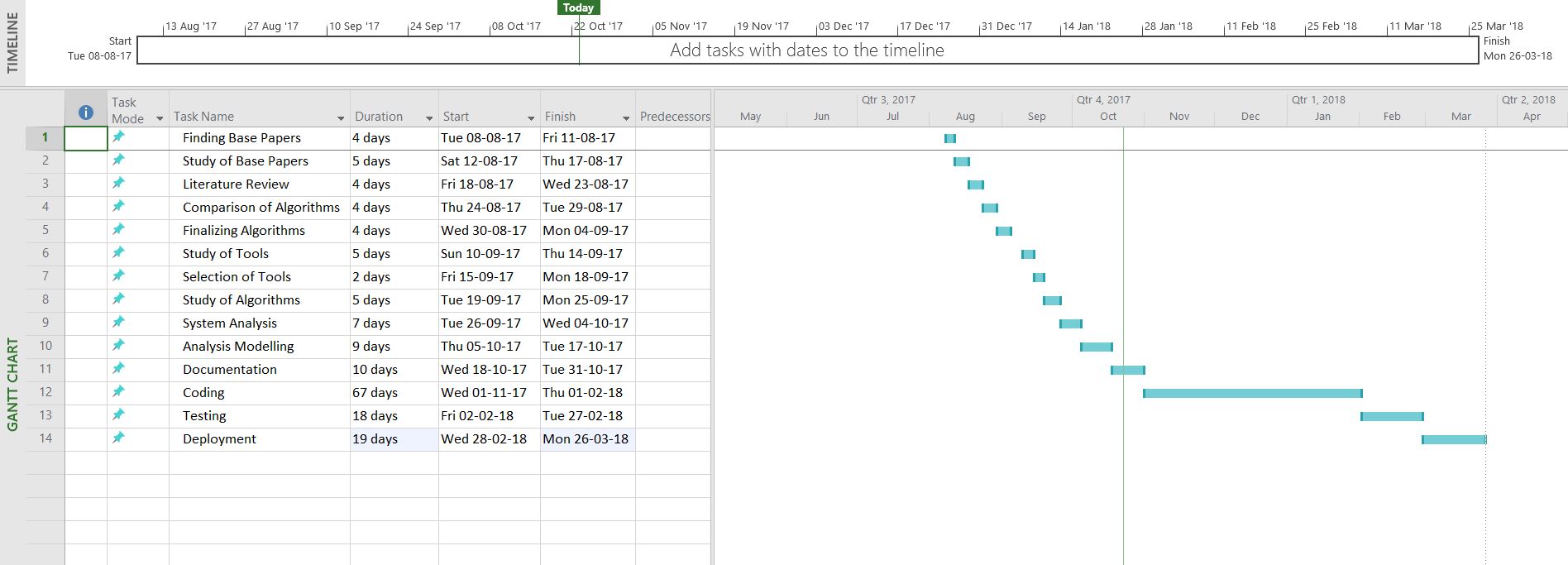






**Fig 4.2.1 Data Flow Diagram**

**4.3 Timeline Chart**

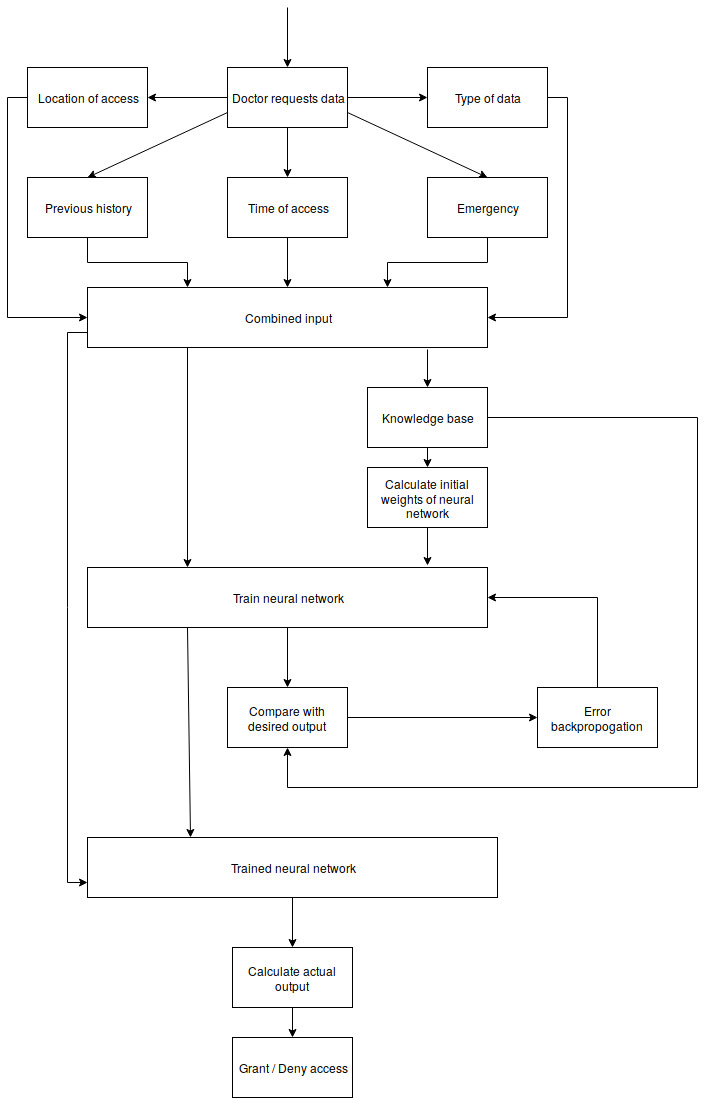
****

**Fig 4.3.1 Timeline Chart**

**Chapter 5**

**DESIGN**

**5.1 Architecture Design**

****

**Fig 5.1.1 System Architecture**

**Chapter 6**

**CONCLUSION AND FUTURE SCOPE**

**Chapter 7**

**ACKNOWLEDGEMENTS**

We would like to express our sincere gratitude to our guide and mentor, Prof. Kriti Srivastava for her guidance, encouragement and gracious support throughout the course of our work, for her expert knowledge in the field that motivated us to work in this area for her faith and belief in us at every stage of this research.

We’re grateful to Dr. Narendra Shekokar, the Head of the Computer Engineering Department, for letting us use the department resources, labs and books.

We are highly obliged to our Principal, Dr. Hari Vasudevan, for availing us with library books and relevant materials and the SVKM management for providing us with well approved and resourceful journals like IEEE and other online materials that helped us finish our project with ease and perfection.

We would also like to thank all our fellow students and staff of the Department of Computer Engineering for their help in the whole process leading to the conceptualization of the project.

Aayush Shah

Pranay Shah

Dipam Vasani