**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 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ . 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.