Health care fraud is an intentional deception used in order to obtain unauthorized benefits (Busch, 2007; NHCAA, 2001).

Health care abuse is produced when either the provider practices are inconsistent with sound fiscal, business or medical practices, and result in an unnecessary cost or in reimbursement of services that are not necessary medically or that fail to meet professionally recognized standards for health care (NHCAA, 2001).

Health insurance systems are either sponsored by governments or managed by the private sector, to share the health care costs.
in those countries (Yang and Hwang, 2005; Opit, 1984).

Processing medical claims is an exhausting manual task that involves analysis, checking, resolution and audit of high volumes of medical claims daily within a limited period of three days from their reception. These control activities are done by few medical experts who have the responsibility of approving, modifying or rejecting the subsidies solicited.

In traditional methods of health care fraud and abuse detection, a few auditors handle thousands of paper healthcare claims. In reality, they have little time for each claim, focusing on certain characteristics of a claim without paying attention to the comprehensive picture of a provider’s behavior (Rashidian et al., 2012). This method is time-consuming and inefficient.

Due to the complexity of the processes within healthcare insurance systems and the large number of participants, it is very difficult to supervise healthcare insurance systems for fraud and abuse.

Detecting service provider’s fraudulent and abuse, needs intensive medical knowledge and currently the task is often conducted by experts that manually review insurance claims and identify suspicious ones. Therefore, an effective fraud detection system is necessary for improving the quality and reducing the cost of health care services.

In order to assure the healthy operation of a healthcare insurance system, fraud and abuse detection mechanisms are imperative. In this work, a healthcare insurance fraud and abuse detection system is proposed. The system is capable of detecting service providers who are practicing inappropriately, such as those who perform more services more than necessary, see their patient more often than warranted or even bill non-rendered services. The development of this system will overcome the limitation of existing approaches.

**Fraudulent and Abusive Acts**

The common fraudulent and abusive behaviour pertaining to the three parties (service providers, insurance subscriber’s and insurance carriers) involved in healthcare insurance services cannot be overemphasized. The service providers fraud and abuse include: billing for services that were never rendered, performing more expensive services and procedures, performing unnecessary medical services solely for the purpose of generating insurance payments, misrepresenting non-covered treatment as necessary medically covered treatment for the purposes of obtaining insurance payments and falsification of patients diagnosis and/or treatment histories. The insurance subscriber’s fraud and abuse include: misrepresenting application for obtaining lower premium rate, falsification of records of employment/eligibility and falsification of medical claims. The insurance carrier’s fraud and abuse include: falsification of reimbursements and benefit/service statements.

**LITERATURE REVIEW**

A study for fraud indicators and rules from the knowledge and experience of human experts to develop a computer-based expert to facilitate the work of insurance carriers were carried out by Herb and Tom (1995).

A neural networks was used by Brockett et al. (1998) to classify fraudulent and non-fraudulent claims for automobile bodily injury in healthcare insurance claims.
The classification algorithm C4.5 was applied for fraud and abuse detection by using the discovered temporal patterns as predictive features. Some data mining approaches which can be used to extract medical knowledge for diagnosis, screening, prognosis, monitoring, therapy support or overall patient management were presented by Lavrac (1999).

Wei et al. (2000) and Hwang et al. (2004) presented a temporal pattern mining algorithms to identify a set of frequent temporal patterns gathering insurance claim instances about pelvic inflammatory disease from regional hospitals in Taiwan.

Tasks performed in support of a data mining project for Health Care Financing Administration (HCFA) such as customer discussions, data extraction and cleaning, transformation of the database, and auditing of the data was described in Sokol et al. (2001).

Shapiro (2002) in his overview applied the merger of neural networks, fuzzy logic and genetic algorithms to the insurance industry. A data mining framework that uses the concept of clinical pathways (or integrated care pathways) was utilized for detecting unknown fraud and abusive cases in a real-world data set gathered from the National Health Insurance (NHI) program in Taiwan (Yang and Hwang, 2005).

A supervised method was used by Liou et al. (2008) to review claims submitted to Taiwan’s National Health Insurance for diabetic outpatient services. The authors selected nine expense-related variables and compared them in two groups of fraudulent and non-fraudulent claims for building the detection models. The input variables were average drug cost, average diagnosis fee, average amount claimed, average days of drug dispense, average medical expenditure per day, average consultation and treatment fees, average drug cost per day, average dispensing service fees and average drug cost per day. They further compared three data mining methods including logistic regressions, neural networks and classification trees for the detection of fraudulent or abusive behavior. Liou et al. (2008) concluded that while all three methods were accurate, the classification tree model performs the best with an overall correct identification rate of 99%.

Shin et al. (2012) carried out a study to identify abuse in 3705 internal medicine outpatient clinics. In their study, they gathered data from practitioner outpatient care claims submitted to a health insurance organization. The authors calculated a risk score for indicating the degree of likelihood of abuse by the providers; and then classified providers using a decision tree. The advantage is that they used a simple definition of anomaly score and extracted 38 features for detecting abuse.

Ekina et al. (2013) applied Bayesian co-clustering methods to identify potential fraudulent providers and beneficiaries who might have perpetrated a “conspiracy fraud”.
METHODOLOGY

Architecture of a health care insurance fraud and abuse detection system (HECIFADES)

The architecture of HECIFADES is presented in Figure 1. The HECIFADES consist of six components namely: claim, augment claim, claim database, profile database, profile updater and Updated Profiles.

![Figure 1: Architecture of a health care insurance fraud and abuse detection system (HECIFADES)](image)

**System Architecture**

(a) **Claim**: A medical claim typically includes data such as date of service, performing provider (as well as billing and referring providers), amount paid, patient ID, service-code, diagnosis codes, etc.

(b) **Augment Approach**: medical claims are evaluated using repeated service and/or pair wise co-occurrence approach to derive an augmented claim record.

(c) **Claim Database**: usually found in the healthcare line of insurance either follows the professional services standard format, the facilities standard format or the dental standard format.

(d) **Profile Database**: includes profiles of providers, clients, and provider/client pairs.

(e) **Profile Updater**: The augmented claim records are combined with the profiles database to update the profiles according to profile updater.

(f) **Updated Profiles**: are sent to a scoring/analysis engine (e.g., a predictive model) so that a decision can be made whether to investigate suspicious activity, either at the claim, client, or provider level.

**Conceptual Flow**

In Figure 2, the provider provides a service to client. A transaction record of the service is submitted as claim to, e.g., a health insurance company for reimbursement. Analysis engine scores the claim to determine the likelihood that the claim is fraudulent. The claim and score are provided to a claims analyst. If
the score is below a certain threshold, the claim is reimbursed or otherwise disposed of without investigation. If the score is above the threshold, the claim, the client, and/or the provider are further investigated for fraudulent activity. Depending on the outcome of the investigation, the claim is either reimbursed or not.

**IMPLEMENTATION AND SYSTEM PERFORMANCE**

In this section the HECIFADES design was implemented based on the following software and the hardware requirements.

**Choice of Software Tools**

Visual Studio 2010 was chosen as the application programming language for the development of the GUI and the front-end applications. Similarly, SQL was chosen as the choice language for our back-end applications. Internet Information System 7 (IIS 7) and Entity Framework 5 are other software requirements that support the running of developed system.

**Choice of Hardware Tools**

This project works fully with the following hardware components: a Pentium® Dual-Core CPU, 256MB RAM, 10GB HDD and at least 1200MHZ processor’s speed.

**Database Model and Structure**

Figure 3 shows the database model and structure while Figure 4 shows the class diagram of the HECIFADES design.
**User Interface**

Interfaces of the HECIFADES design shows the home page, setup, operation, report, contact and help menu respectively.

The hospital setup interface as shown in Figure 5, is where the hospital name, address, city id, phone number are been created while Figure 6 shows the hospital standard service rate setup where the hospital name, the service render and the rate for each services are been created.
Figure 7 shows the provider setup where the provider number and name are been created while Figure 8 shows the insurer setup interface where the insurer surname, other name, sex, address, state, city, date of birth and provider name are entered.

Figure 7: Provider Setup

Figure 8: Insurer Setup

Data about the service render by the hospital are created under the operation menu of the HECIFADES as shown in Figure 9. The data entered include hospital name, insurer id, types of service render, charges, date in which the service was rendered, provider name and medical id number.
The report generated on services provider, hospital standard rate, hospital services, list of insurers and claims are shown in Figures 10, 11, 12, 13 and 14 respectively.

Figure 9: Operation - Hospital Services

Figure 10: Report on Services Provider
Figure 11: Report on Hospital Standard Rate

Figure 12: Report on Hospital Services

Figure 13: Report on List of Insurers
CONCLUSION
As more people commit health care fraud, the more complex types of fraud are being created. This study has presented architecture for health care insurance fraud and abuse detection. The new design targeted at minimizing fraudulent activities and abuse of claims found in health care insurance.

The system detects, prevents and decreases improper payments associated with fraud, and also improves the ability to identify and mitigate potential payment risks and program vulnerabilities.

The best way to prevent health care fraud and abuse is to address these issues now, and in order to do so it is expected that in future mobile devices (e.g., smartphone, tablet PC’s, etc) which are increasingly becoming an essential part of human life which is the most effective and convenient communication tools will be incorporated into the health care insurance system. It might also be necessary that cloud technology will be adapted to health care insurance.

REFERENCES


(Manuscript received 7th April, 2014; accepted 5th March, 2015).