

Chapter 12

Supervised Learning Methods for Fraud Detection in Healthcare Insurance

Prerna Dua and Sonali Bais

Abstract Fraud in the healthcare system is a major problem whose rampant growth has deeply affected the US government. In addition to financial losses incurred due to this fraud, patients who genuinely need medical care suffer because of unavailability of services which in turn incur due lack of funds. Healthcare fraud is committed in different ways at different levels, making the fraud detection process more challenging. The data used for detecting healthcare fraud, primarily provided by insurance companies, is massive, making it impossible to audit manually for fraudulent behavior. Data-mining and Machine-Learning techniques holds the promise to provide sophisticated tools for the analysis of fraudulent patterns in these vast health insurance databases. Among the data mining methodologies, supervised classification has emerged as a key step in understanding the activity of fraudulent and non-fraudulent transactions as they can be trained and adjusted to detect complex and growing fraud schemes. This chapter provides a comprehensive survey of those data-mining fraud detection models based on supervised machine-learning techniques for fraud detection in healthcare.

Keywords Healthcare fraud · Fraud detection · Supervised methods · Unsupervised methods

P. Dua (✉)

Department of Health Informatics and Information Management, Louisiana Tech University, Ruston, LA, USA
e-mail: prerna@latech.edu

P. Dua

School of Biological Sciences, Louisiana Tech University, Ruston, LA, USA

S. Bais

Department of Computer Science, Louisiana Tech University, Ruston, LA, USA

12.1 Introduction

Healthcare fraud, a severe and challenging problem faced by medical providers, can be defined as an offense committed by an individual or group of individuals who place false medical claims for services that have never been used to gain unauthorized financial benefits. According to the data provided by Centers for Medicare and Medicaid Services (CMS), the United States spent overall \$2.5 trillion after healthcare during the 2009 fiscal year. This expenditure signifies \$8,086 per person or 17.6 % of the Gross Domestic Product (GDP), increased from 16.6 % in the 2008 fiscal year [1]. It is estimated that over five billion health insurance claims were paid that year [2], and some of them were fraudulent. Even though these fraudulent claims composed only a small portion of the claims, they carried a very high cost value. It is predicted by CMS that the healthcare expenditure may increase up to \$4.14 trillion by 2016, signifying 19.6 % of the GDP [3]. Information provided by National Health Care Anti-Fraud Association (NHCAA) shows that approximately \$60 billion, or 3 %, of healthcare spending has been expended on healthcare fraud [4]. This quantity of money is more than the GDP of 120 countries including Kenya, Ecuador, and Iceland [5]. If steps against healthcare fraud are not taken, such expenses can affect quality of life and national economies. The Federal Bureau of Investigation (FBI) approximates that between \$70 and \$234 billion are stolen from US citizens in healthcare annually due to fraud in the healthcare system [6]. Even if the financial loss is disregarded, healthcare fraud can hamper the healthcare system of the US from offering good services and care to patients. Hence, the efficient detection of fraud is vital, as it allows for quality enhancements and the lowering the expenditures to healthcare services.

To detect fraud within the healthcare system, the process of auditing is followed by investigation. If accounts are carefully audited, it is possible to identify suspicious policy holders and providers. Ideally, all claims should be audited carefully and individually. However, it is impossible to audit all claims by any practical means as these form huge piles of data involving sorting operations and complex computation [23]. Moreover, it is difficult to audit service providers without clues as to what auditors should be looking for. A sensible approach is to make short lists for scrutiny and audit patients and providers based on these lists. A variety of analytic techniques can be used to compile audit short lists. Fraudulent claims frequently build into patterns that can be perceived using predictive models.

12.2 Types of Fraud in Health Care

Healthcare fraud can be divided into four types: (2.1) medical service providers, (2.2) medical resource providers, (2.3) insurance policy holders, and (2.4) insurance policy providers. Figure 12.1 demonstrates the overview of fraudulent activities found in healthcare.

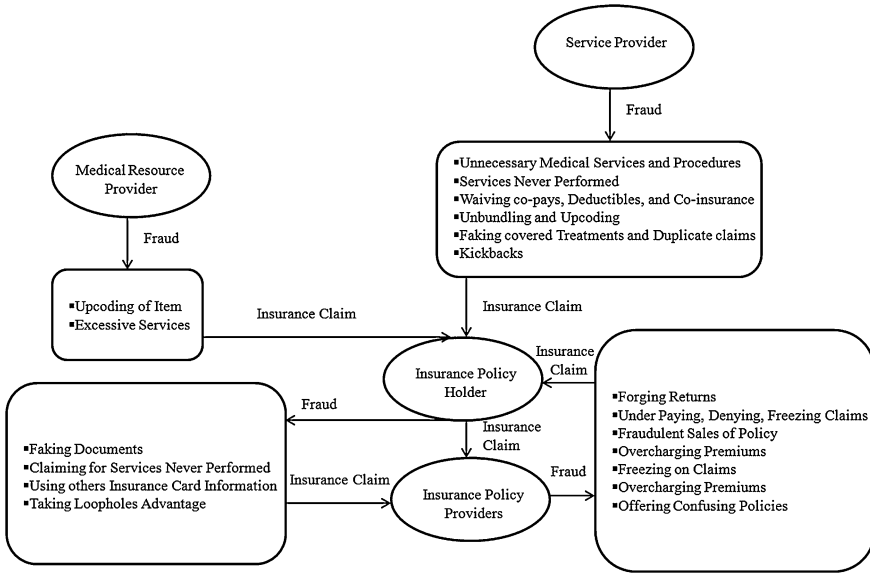


Fig. 12.1 Types of fraud in healthcare system

12.2.1 Medical Service Providers

Medical service providers can be hospitals, doctors, nurses, radiologists and other laboratory service providers, and ambulance companies. Activities involving Medical Service Providers may include:

- Forging a patient diagnoses to rationalize services and procedures that are not medically required [7],
- Billing for services which were never performed using authentic patient information, aiding in identity theft, or modifying claims with extra charges for the services or procedures that were never performed [7],
- Billing the insurance companies more by waiving patient co-pays, deductibles, or co-insurance [7],
- Billing for unnecessary services or procedures, such as daily checkups rather than monthly ones, only to create insurance payments [3, 7],
- Billing each phase of a process as if it were a different process, known as unbundling [7], for example billing tests within test “sessions” as if they were independent sessions [3],
- Billing for expensive services instead of billing for low cost services or procedures which were performed, known as upcoding or coding a patient diagnosis to a more critical and expensive charge and applying charges with false CPT codes, such as charging a 30 min group therapy as a 50 min personal therapy [3, 7],

- Billing for duplicate claims for a particular service or an item, in which the provider changes part of the claim for example the date on which the service was provided to fool the health insurance company into paying for a single service or item twice [3], and
- Accepting illegal kickback schemes in which healthcare providers trade money or something of value for patient referrals, for those healthcare services that can be paid by Medicaid or Medicare [3].

An example of fraud by healthcare providers is the Federal indictment of the Benitez brothers, owners of a chain of medical clinics in the Miami area that treated HIV-infected Medicare beneficiaries. The Benitez brothers were accused of submitting fraudulent insurance claims that cost Medicare approximately \$110 million. For these claims, they performed unnecessary procedures on the patients. They also offered kickbacks to the patients in exchange of their Medicare information, which was used to submit false claims to Medicare for compensation [9].

12.2.2 Medical Resource Providers

Medical resource providers can be pharmaceutical companies, medical equipment companies that supply products like wheelchairs, walkers, specialized hospital beds and medical kits. Activities involving Medical Resource Providers may include:

- Medical equipments manufacturer put forward complimentary products before the patients. Then they charge patients insurance company for the free product which is either not required or was never delivered [26],
- Sometimes needless or false tests are provided to the individuals at shopping malls, old-age homes, and health clubs and billed to insurance companies or Medicare [26],
- Medical resource providers also bills individuals insurance companies for resources never been delivered by altering the bills or submitting forge ones [26],
- The common target of medical resource providers are the senior citizens who are offered free products by these providers in exchange of their Medicare insurance number. In order to get paid by the Medicare, the doctor is required to sign a form certifying that the equipment is required by the patient as a part of medical treatment. To achieve this, medical resource provider either bribes the corrupt doctors or provide fake signature for billing Medicare for the equipments which are either not delivered or not required [26],
- Up-coding items, for instance sending a patient an essential, manually pushed wheelchair but billing the patient's health insurance plan for a power-driven wheelchair [3] and
- Providing excessive services that the patient does not need, for example, delivering and billing for 40 injury care kits per week for a patient in hospital who only needs a change of dressings once per day [3].

In August 2004, Pfizer employees were convicted of being involved in illegal kickback schemes and off-label endorsements. In this case, a Pfizer northeast regional manager directed around a hundred sales employees to market a painkiller called Bextra, which was prohibited by FDA. The FDA had approved Bextra for distribution in U.S., but citing augmented occurrences of strokes, heart attacks, and severe skin reaction to drug, petitioned Pfizer to withdraw this drug from U.S. market [3].

12.2.3 Insurance Policy Holders

Insurance policy holders consist of individuals and groups who carry insurance policies, including both patients and employers of patients. Activities involving Insurance Policy Holders may include:

- Faking documents related to employment or eligibility to obtain low premium rates and good benefits [20],
 - Filing claims for medical services or procedures which never took place [20],
 - Claiming insurance benefits illegally using someone else's coverage or insurance card information [20], and
 - Taking advantage of insurance benefits by finding loopholes in the policy.
- In 2007, a fraud case was committed by falsely filing a life insurance claim. The fraudulent owner faked his own death in a canoeing accident and lived a secret life in his house for five years. His family claimed the money from the insurance company, so that he and his family could start a hotel business for canoeing holidays in Panama [27].

12.2.4 Insurance Policy Providers

Insurance policy providers are the entities that pay medical expenses to a policy holder in return for their monthly premiums. Insurance policy providers can be private insurance companies or government administrated healthcare departments including agents and brokers. Very little research has been conducted regarding fraud committed by insurance policy providers as most insurance fraud data are provided by the providers. It is estimated that around \$85 billion are lost yearly due fraud committed by insurance companies [28]. Activities involving Insurance Policy Providers may include:

- Forging returns and benefit/service statements by under paying claims [20],
- Insurance company wrongfully denies valid claims to try to discourage the policy holder and hoping that the patient will eventually give up [28],
- Freezing claims without investigating the merits of claims [28],

- Overcharging the premiums from policy holders by misinterpreting the information to the client and making them pay for those coverage which they don't actually have [28],
- Creating confusing policies that mislead policyholders on coverage issues [28], and
- Making fraudulent sales of fake policies which are of no use to the policy holders and are mainly intended to get high premium from them.

During September 2009, an individual lost her health coverage by Blue Cross Insurance Company because the company discovered her pre-existing condition. This company terminated her coverage because she never mentioned her pre-existing condition, which she herself was unaware of initially. Hence, the company abruptly cancelled her coverage after she was diagnosed of thyroid disorder and fluid in the heart leaving her in a debt of \$25,000 in medical bills [40].

Among these four types of fraud discussed above, the service providers alone commit the majority of the fraud. Although most service providers are trustworthy, those few dishonest service providers commit fraud and cause the loss of millions of dollars to the healthcare system. In some cases, more than one of the above types is involved in committing healthcare fraud. Detecting fraud in such a hybrid cases can be complex and challenging [20]. Hence, it is urgent that researchers find effective ways to discover patterns and relationships in data that may be used to make a valid prediction about fraudulent claims. Due to this pressing need, high end data mining and machine learning techniques holds a promise to provide sophisticated tools to identify possible predictors that characterize the fraudulent behaviors based on the historical data [20].

12.3 Data Mining for the Fraud Detection in Healthcare

Data mining is a popular means for detecting fraud and abuse in the healthcare system. The vast quantities of data produced by healthcare insurance companies are difficult to process and evaluate using conventional methods. Data mining provides the techniques and expertise to convert these heaps of data into the useful collection of facts for decision making [8]. This kind of analysis is becoming increasingly important, as financial pressure has increased the requirement for healthcare industries to construct judgments based on the study of fiscal and clinical data. Information and analyses obtained through data mining can improve operating efficiency, decrease costs, and increase profits while preserving a high-level of care [30].

Additional reasons behind the increasing popularity of data mining include the use of fee and categorization systems. For example as an outcome of the Balanced Budget Act of 1997, CMS have to employ a potential fee system supported categorizing patients into case-mix clusters, with the help of empirical proof that supplies utilized within every case-mix cluster are comparatively constant. CMS

has utilized data-mining techniques to build a potential reimbursement system for inpatient treatment [31].

The data-mining applications generally establish norms for detecting fraud and abuse. Then, these applications identify unusual patterns of claims by clinics, laboratories, and physicians. Along with other details, these data-mining applications can provide information about out of place referrals, prescriptions, medical claims and fraudulent insurance claims. For example, the Texas Medicaid Fraud and Abuse Detection System collected a great deal of data produced by millions of treatment courses, operations, and prescriptions to recognize abnormal behaviors and discover fraud. It recovered \$2.2 million and successfully recognized 1,400 suspects for inquiry in 1998. This result is impressive considering that it was obtained after only a few months of use [32]. Due to this achievement, the Texas system was awarded a national prize for this innovative utilization of the expertise.

Data-mining techniques can be categorized into supervised methods and unsupervised methods.

Supervised Data-Mining Methods

Supervised machine-learning techniques consists of algorithms that reason from outwardly given instances to construct universal theorems which then predict upcoming instances. Supervised machine learning is used to construct a brief model of the allotment of class labels which refer to predictor features. Then, the testing instances are assigned class labels based on the resulting classifier, in which the predictor feature significance is known, but the value of the class label is unknown [22].

In this context, the weightage is given to those healthcare fraud detection models which implement supervised machine-learning techniques. Supervised methods like Neural Networks [13, 16], Association Rules [14, 15, 17], Genetic Algorithms [10], Fuzzy Bayesian classifier [11], Logical Regression methods [16, 21], Bayesian Networks [12, 19], KNN classifiers [10], and Classification Trees [16] have been used by researchers to detect fraud in the healthcare system.

Unsupervised Data-Mining Methods

Unlike supervised data-mining methods, unsupervised methods do not get any objective output or benefits from their surroundings. Although it is difficult to visualize how a machine can be trained without any response from its surroundings, these methods work well. It is very likely to build a proper model for unsupervised learning methods supported on the idea that the mechanism's aim is to use input characterization to foresee prospective input, effectively communicating the input to another mechanism, decision making, and so on. It can be said that unsupervised learning can find patterns in a data which can also be unstructured noise. Clustering and dimensionality reduction are the classic examples of unsupervised learning [22].

The benefit of using supervised techniques over unsupervised is that once the classifier has been trained, it can be easily utilized on any same kind of datasets [41]

which makes it a best choice for a fraud detection program which involves screening and monitoring. In this chapter, we only consider supervised machine-learning techniques and provide an in-depth survey of their application in detecting fraud in the healthcare system.

12.3.1 Data Sources

The raw data for detecting healthcare fraud is obtained from insurance companies which explain the reason for decreased research performed in detecting fraud committed by insurance companies. The insurance company designation can include government administered healthcare departments like Medicare or private insurance companies [20]. The following data sources that have been developed and published in various sources through supervised machine-learning techniques have helped shape the healthcare detection fraud model:

- Health Insurance Commission (HIC) of Australia [10, 14],
- Medicare Australia's Enterprise Data Warehouse, Prism [17],
- National Health Insurance (NHI) Program in Taiwan [15],
- Taiwan's National Health Insurance (NHI) System [16],
- ISAPRE (A private health insurance company in Chile) [13, 21] and Banmedica S.A. [13], and
- Taipei Health Insurance Bureau [11].

Generally, the raw data provided from the above sources consists of insurance claims. The content of insurance claims is related to the service provider and insurance subscriber. These databases contain rich features that are helpful to the fraud detection model in identifying fraudulent patterns of behavior by insurance holders and healthcare service providers. It is possible to gain an overall perspective of both insurance holder and healthcare service provider behavior over time using this information. This overall perspective helps detect fraud committed by these entities [20].

12.3.2 Algorithms

This section provides a discussion on the algorithms based on supervised models for detecting fraud in healthcare.

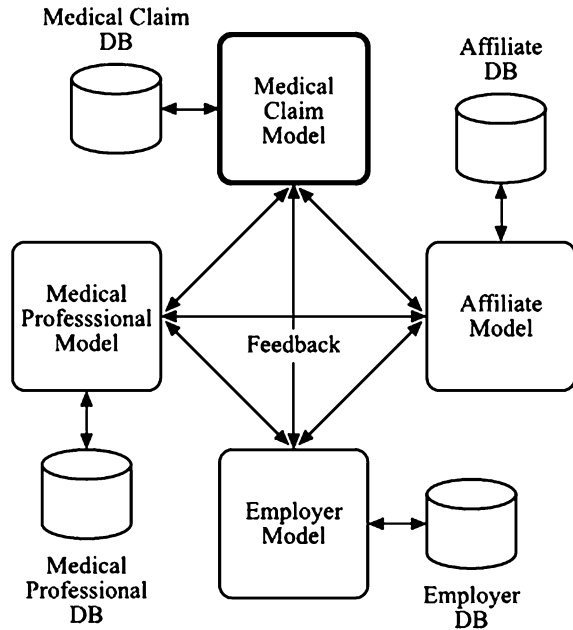
12.3.2.1 Neural Network

A neural network imitates just like the brain of human beings in order to forecast and classify data. A neural network is composed of a set of connected, simulated neurons, few of which get scalar information from other neurons and convert that information into a singular output signal. All the inter associations are weighted among them and customization is done to these weights while the neural network runs on training data. A neural network comprises of a *layered, feedforward, completely connected* network of artificial neuron, or nodes. The term feedforward suggests that the data flows in one direction from input to the output layer. A classic neural network for classifying data can contain two or more layers, although most contain three layers such as, input, hidden and an output layer [42]. Multilayer Perceptron (MLP) is a classic example of a multilayer feed-forward network with one or more hidden layers between the input and output layer. Neurons in the hidden layer obtains a weighted summation of the input variable, and convert that total to a signal in the form of output with the help of a threshold function such as a sigmoid or step function. The weighted summation, which is obtained from the hidden layer, is given to the single node of an output layer and is transformed into a classification signal. Neural networks set up an association amid the input and yielded data, and are efficient with noisy data [16].

Ortega et al. [13] proposed a fraud detection system which utilizes committee of MLP networks for each entity (such as, medical claims, affiliates, medical professionals and employers) involved in the fraud, in a Chilean private health insurance company. Figure 12.2 demonstrates the four sub-models which are a working group of neural networks for all entities. The inputs are pre-calculated attribute vectors that represent the specific exploitation and fraud sub-problem. Whenever a medical claim is obtained by the ISAPRE (a private pre-paid health insurance plan) system, the output of each committee conveys a predictive value. These values provide supplementary inputs to the sub-models, offering a response technique for merging the diverse outcomes. Models are assessed at fixed time intervals as per the predetermined agenda. The model representing medical claim is implemented every day to process inward entries while the other models are implemented once in a month. Every sub-model is trained again on monthly basis. A data renewal process has been defined for keeping the training models descriptive for historical and new fraud patterns. With the help of experts new training samples are selected and thoroughly classified. Then a subclass of equivalent amounts of normal and fraud conditions are chosen and included to the training dataset. Hence, the model retains knowledge of new kinds of fraud, and is able to prevent these emerging categories [13].

The drawback of using a neural network is that it cannot identify the importance of individual variables. To overcome this problem, Liou et al. [16] used neural nets in detecting fraud and claim abuse based on diabetic outpatient services, which helped perform sensitivity analysis among variables. The authors then presented those variables which were most significant for classification. The end result specified the order of every variable's relational significance in categorizing the

Fig. 12.2 Fraud detection scheme by using sub-models with feedback connections



data like averages of different types of claims like dispensary visiting service fees, diagnosis and prognosis fess, daily medical expenses, cost of medicine per patient, cost of medicine per patient per day, session and disease curing fees, healthcare expenditure, and insurance claimed. The results of applying neural network algorithm to the whole test sample and normal samples were 95.73 and 91.47 % respectively. This results further showed that the neural network model had an error rate of 9 % for classifying normal providers [16].

12.3.2.2 Bayesian Belief Network

Bayesian Belief Networks are dense networks of probabilities which capture information of the probabilistic relationships between variables as well as the past information about the relationships. This technique is very beneficial in those situations where some information is previously known and the incoming data is uncertain. These networks also give reliable semantics for characterizing effects and causes through an intuitive graphical representation. Due to these reasons it is used widely in those domains where automated reasoning is required [43].

In the research conducted by Ormerod et al. [12], a Mass Detection Tool (MDT) based on Bayesian Belief Network has been developed to detect healthcare fraud. This tool offers a real-time response as to the likelihood of several methods of fraud. It also recommends those unfamiliar indicators that can affect fraud possibilities and helps the claims handler improve online decision making.

Figure 12.3 shows the MDT framework, which helps link indicators to explicit types of fraud through a node-link topology. To forecast possible types of fraud for permutations of pointers, the network uses stored probability data. The most important advantage of the Bayesian Belief Networks is automated information updating. This updating is accomplished by the MDT in following three ways [12].

1. With the help of case result information from the other tool called Suspicion Building Tool (SBT), the provisional and previous possibilities of pointers and their associated fraud types are processed.
2. The claims handler classifies unexpected anomalies which uses the taxonomy educed in the ethnography. Then, each abnormality takes over the base rate possibility of identified fraud pointers in the similar subcategory. The information of prospective anomaly helps in reweighting the Bayesian Belief Networks, permitting the set of connections to acquire an explanation of unanticipated information as it arises.
3. A trace of anomalies and their results is developed over each instance. If an anomaly achieves better predictive power than a known pointer, then one is substituted by the other. Bayesian Belief Networks topology is updated with the help of SBT argumentation engine's output.

Hence, by this means the entire classification method maintains pace with the evolving types.

IBM teamed up with fraud investigation specialists and healthcare industry experts to develop a system to help detect fraud in insurance companies, health management organizations (HMOs) and for risk-bearing health care professionals. The system is known as Fraud and Abuse Management System (FAMS) and uses Fuzzy modeling along with decision support techniques for detection, investigation, prevention, and settlement [24]. Used with the Fuzzy modeling system, FAMS assigns a score to those providers who deviate from the normal behavior of their peer group. It has over 650 standard, individual behavior patterns such as percentage of specialty diagnoses and the average number of procedures per visit. For creating an analysis model, users choose and link behavior patterns suitable to the peer group they want to inspect from a library of functional objects. This model consists of around 25–30 behaviors. FAMS, used with analyses of claims data, calculates values for each provider in the model. Then, each value is allocated a score between 0 and 1,000 depending on the degree of deviation from the respective peer group norms. The score is higher if the deviation is greater. For scoring the behavior of each provider, FAMS uses Fuzzy membership functions. The values for each behavior pattern for all providers in the peer group are calculated, and the allocation of these values is analyzed by the system. Only providers are assigned scores with values greater than the peer group median. The investigation priority list is compiled of providers having the highest scores. By using the FAMS analysis tools, the suspected providers can be checked for fraudulent behavior [24]. The overall working of FAM is shown in Fig. 12.4.

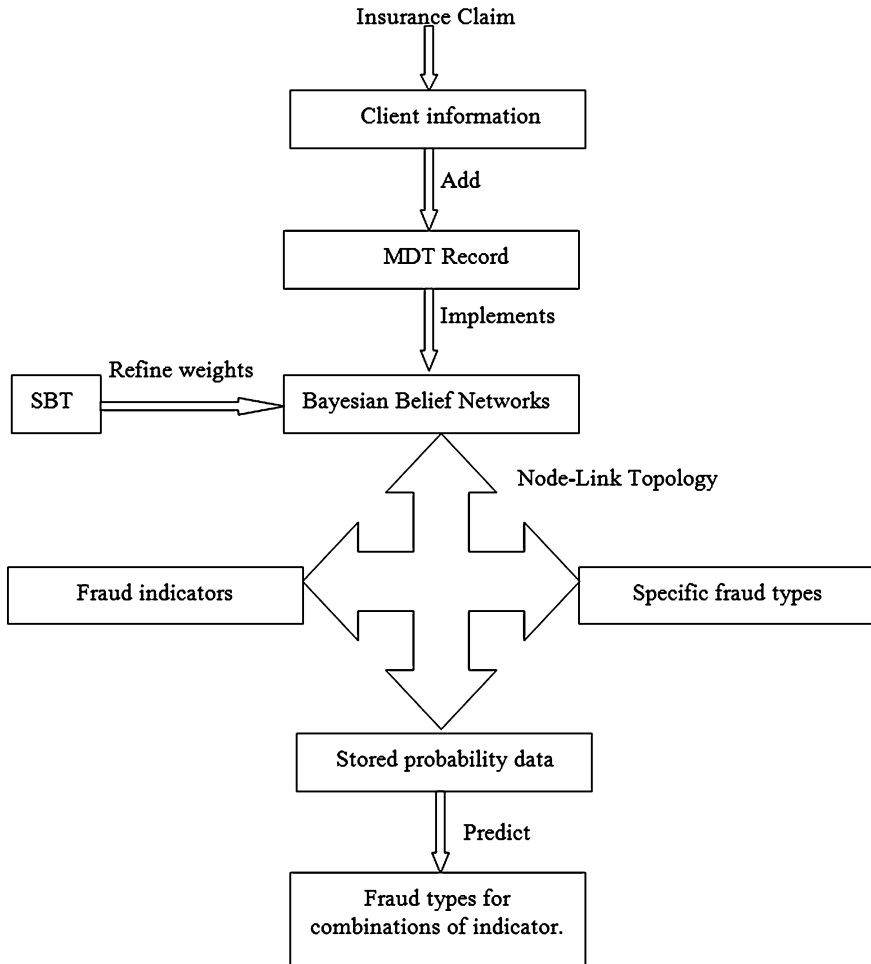


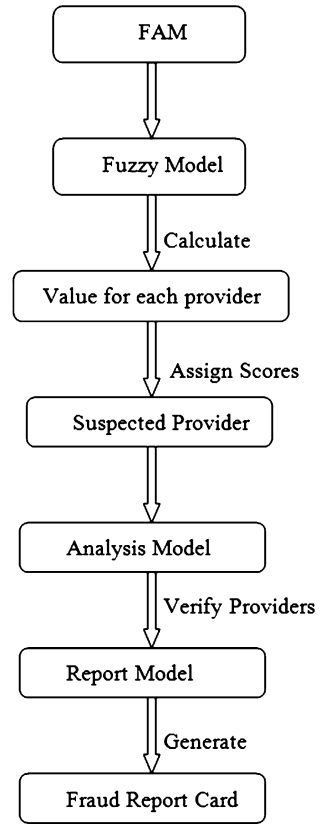
Fig. 12.3 Bayesian belief networks fuzzy modeling

12.3.2.3 Fuzzy Bayesian Classifier

Bayesian classifier is an important data mining technique, which can efficiently achieve optimal results when probability distribution is provided. Bayes rules can be used to computed the posterior from the likelihood and the prior, as the latter two is usually easy to be computed from a probability model [29].

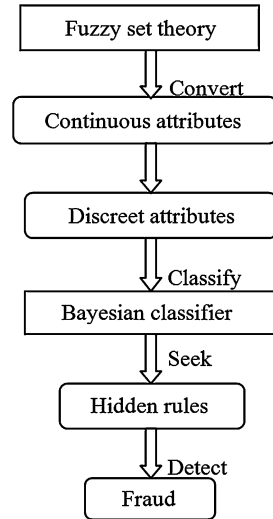
Chan et al. [11] developed a new Fuzzy Bayesian classifier to audit health insurance fee data. Based on Bayesian inference, the Bayesian classifier acquires every attribute influencing the classification outcome. The Bayesian classifier classifies the case set more visibly by having an excellent control and understanding in the interpretation of the results. However, Bayesian inference requires

Fig. 12.4 FAMS fraud detection system



putting together diverse probability distributions while handling continuous attributes. Computation in this scenario will be very complex. This complexity can be overridden when this method is used along with Fuzzy set theory. With this hybrid method, it is possible to convert continuous attributes into discrete attributes. For their experiments, Chan et al. used 800 records containing health insurance fee data. Of these records, 166 were fraudulent and 634 were normal. They used three methods (80/20, 70/30, 60/40) to divide the training and testing datasets. Then, the training data was fed to Bayesian classifier for training the classification rules, and the remaining, testing, dataset was classified. The values of sensitivity, specificity, and accuracy were computed for all three methods. It was found that the quality of the proportion 80/20 was best, as it had the highest sensitivity (0.639), the highest specificity (0.968), and the highest accuracy (0.894), respectively. It is observed that the overall accuracy of the classifier is good, but sensitivity is slightly less than required. The reason for this low sensitivity is that the attributes chosen for detecting the fraud from the health insurance fee data are not adequately represented. Figure 12.5 demonstrates the overview of Fuzzy Bayesian classifier.

Fig. 12.5 Fuzzy bayesian classifier data mining model

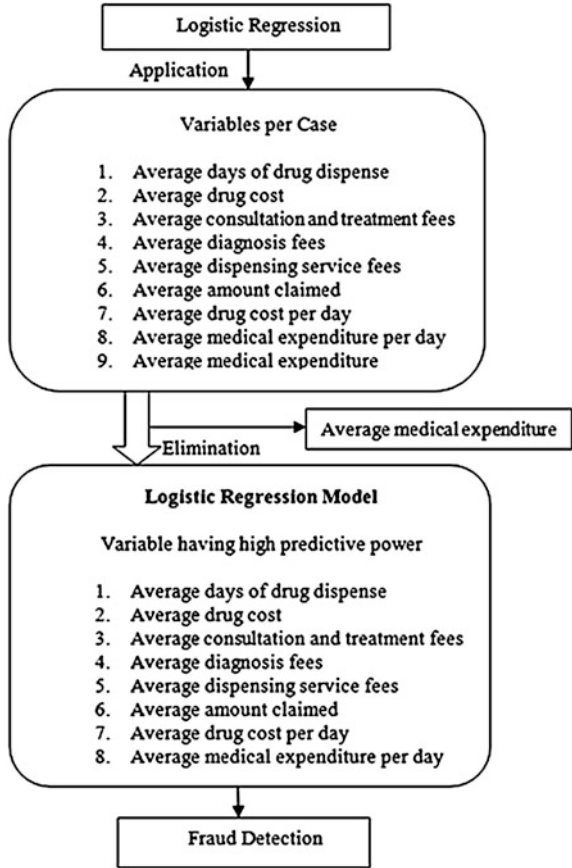


12.3.2.4 Logistic Regression

Logistic Regression is a technique that is nonlinear and is used to represent the variables that are binary dependent. There can be only two values for classification variables, a value that signifies success or a value that signifies failure. The advantage of the logistic regression function is that it can be easily interpreted. The logistic regression technique was used by Liou et al. [16] to detect fraudulent and normal hospitals. In their experiments, they defined that claim was related with the value of zero if it was a regular claim and with the value of one if it was an irregular claim. As shown in Fig. 12.6 nine expense related variables were selected for the detection model. Logical regression was executed on all of these models independently to recognize the most efficient factors. Eight out of nine variables were observed to contain predictive power. The variable that was left out was the average medical expenditure. These eight detectors were then employed to form a complete logistic regression model. The model's successive rate was 100 % for detecting those hospitals involved in fraudulent activities. Along with this rate, the model also had an 84.6 % identification rate for normal hospitals. This rate shows that the logistic regression model has error rate of 15 % in classifying normal providers. The entire sample had a 92.2 % correct identification rate.

In another case, the logistic binomial regression technique was utilized by Table et al. [21] in detecting the health care fraud. Dependent, dichotomous variables were used to represent either fraudulent or non-fraudulent values. Four variables were selected from the database. They were (1) number of days of sick leave granted by the attending physician, (2) amount to be paid for granted days of sick leave, (3) classification of disease by whether it was diagnosable (true) or not (false), and (4) history of health insurance reimbursement claims. With these

Fig. 12.6 Logistic regression model for detection of fraudulent claims



variables, a binomial logistic regression model was developed with the help of following equation.

$$PFLMC = \beta_0 + (\beta_1 D) + (\beta_2 M) + (\beta_3 C) + (\beta_4 N),$$

where $PFLMC$ = probability of fraud, β_0 = constant regression, β_{1-4} = specific beta coefficients for each of the four independent variables, D = days of medical leave granted by the treating physician, M = amount to be paid in cash for sick days, C = diagnostic classification, and N = number of sick leave accumulated by an individual.

The variables with diagnostic classification and existence of multiple requests for leave of absence of an individual had strong predictive power, while variables demonstrating leave and total pay showed a marginally significant predictive power. The most characteristic features of these fraudulent behaviors were the existence of multiple requests for leaves of absence for one person and difficult diagnostic testing. If the number of sick days increase, then it was more likely that

fraud is taking place. The model performed well with sensitivity (99.71 %) and specificity (99.86 %). The positive predictive value, or percentage of fault detected correctly by the model, was around 98.59 %, and the negative predictive value, or the percentage of non-fraudulent cases detected correctly by the model, was 99.97 %.

12.3.2.5 Classification Tree

Classification trees are built with the help of rules, training sets, samples in the dataset, and can directly be applied in the form of a simple detection algorithm. The sequence of classification rules is signified by the “every probable pathway from the root node to a leaf node” technique as shown in Fig. 12.7. Liou et al. [16] were able to detect 100 % of fraudulent hospitals using a classification rule. The correct identification rate for the fraudulent hospitals was 99.30 % for the entire dataset and 98.73 % for the normal hospitals. This results showed 1 % fault rate in categorizing normal providers. Sequence of some classification rules for fraud detection can be seen in Fig. 12.7 and more rules depending on the requirement of the system can be added further.

12.3.2.6 Genetic Algorithm

Genetic algorithm (GA) is a search heuristic which is based on the process of natural selection and genetics. GA's not only perform better than other traditional methods in the majority of the problem link but also offers different methods in the majority of the problem link. GA has the ability of finding optimal parameters for the real world problems which is quite hard in the case of traditional methods [44].

The GA is considered as an ideal technique to solve optimization problems. Its application gets better matches amid when used with examination by expert consultants and the classifications of a system. He et al. [10] used the GA for selection, crossover, mutation, and cost functions to discover the best possible weighting of attributes utilized to categorize the practice profile of general practitioners. They utilized a validation dataset to optimize the weight. In this research, the GA proved to be effective, as each run required only a 2,000 generation, getting the preferred agreement rate for the given validation dataset. Figure 12.8 demonstrates the working of GA.

12.3.2.7 K Nearest Neighbor

K-nearest neighbor (KNN) is an extensively employed profile-matching method that establishes categorizations of every case on its nearest neighbors by applying several decision rules. As shown in Fig. 12.9, He et al. [10] used the weights of the features determined by genetic algorithm in the KNN technique to recognize the

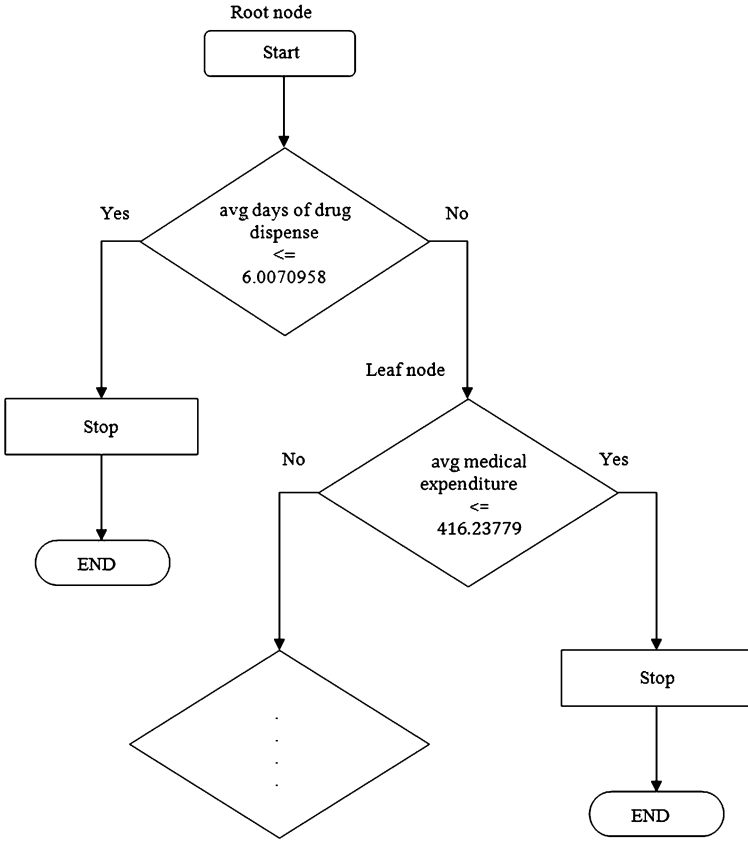


Fig. 12.7 A snapshot of rules involved in a classification tree

Fig. 12.8 Working of genetic algorithm

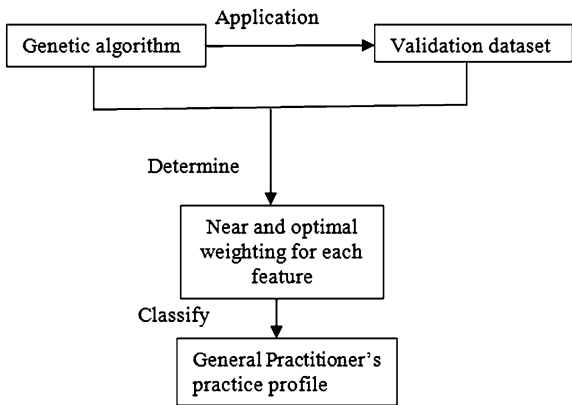
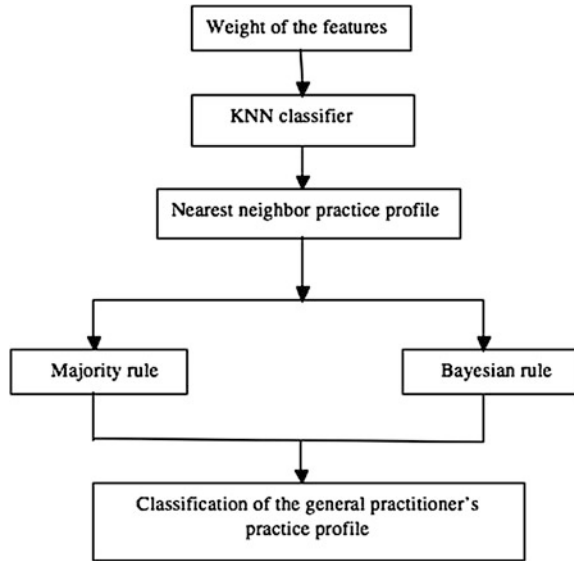


Fig. 12.9 Application of KNN classifier



nearest neighbor profile of the general practitioners. Then, the majority and Bayesian rules were used to classify the general practitioner profiles. The KNN classifier was trained with the help of the nearest neighbor examples provided by the profiles in the training dataset. The trained classifier was then tested on the test dataset. The KNN classifier, along with the genetic algorithm for optimizing the weights, uses Euclidean distance to improve the results of classification. The statistic employed to measure the efficiency of the KNN classifier and variants was the synchronization rate that is basically the percentage of synchronization amid the categorizations of both the KNN classifier and that of expert consultants partitioned by the number of cases in the dataset. Using the majority and the Bayesian rule along with the KNN classifier helped achieve the high agreement rate in this scenario.

12.3.2.8 Association Rules

The use of association rules is one of the many data-mining techniques applied to discover corresponding relationships and remarkable association among a huge array of data items. Association rules demonstrate feature value states which take place together in a known dataset repeatedly. This type of information is given in *if-then* statements by these rules. These rules are determined from the data set provided. Unlike, if-then rules that have logical characteristics, association rules have probabilistic attributes. Along with the precursor that is the 'if' part and the descendent 'then' part, an association rule has an additional two numbers which state the measure of improbability regarding the rule. In association analysis, the

precursor and descendent are itemsets which do not have mutual entries. The first number is identified as support for the rule which is basically the number of transactions that contain every item in the precursor and descendent section of the rule. Sometimes support is denoted as a fraction of the sum of entire records in the database. The second number is identified as the confidence of the rule which is the relative amount of the number of transaction that contain every items in the precursor as well as the descendent to the number of transactions that contain every items in the descendent [18].

Association rules have been widely used [14, 15, 17] for detecting fraud committed by medical service providers in the healthcare system. Until recently, positive association rules have been used to discover frequent patterns. However, the use of negative association rules [17] has proved effective in detecting fraud in the healthcare system. Shan et al. [17] identified around 215 association rules from the dataset, which consisted of 23 positive and 192 negative association rules for detecting inappropriate billing by specialists. The negative association rules outnumbered positive association rules because negative rules of both the presence and absence of the item were found, while only presence of item was considered for positive rules. In addition, negative rules were stronger than positive rules in terms of confidence; minimum confidence of negative association rules was 95.95 %, while for positive rules it was 80.25 %. The common patterns corresponding to negative rules were considered reliable, with the billing regulations enclosed under the Medicare Benefit Schedule. It was found that negative rules proved to be more intuitive and useful for locating violations than positive rules were. The violations of negative rules included billing items that were not generally billed by the majority of specialists. Those specialists who were found to be violating these rules repeatedly were marked unusual from their peers. Among 192 negative rules, 30 were found to have confidence value of 1.0 and were considered unimportant for fulfillment. Hence, these rules were removed. The remaining 162 rules were divided into three groups based on the probability of improper billing by a subject matter expert. A high rating suggested that a rule was crucial, and if this rule was broken then there was a high probability of inappropriate billing to Medicare Australia. On the other hand, a low rating suggested that if a rule was broken, then this behavior may indicate inappropriate billing or another valid explanation for billing may also exist. Hence, it was determined that a low rating may not be powerful for finding inappropriate billing information, but may be helpful in obtaining useful information on identifying specialists that had related compliance activities. It was proven from the experiment that more than half of the rules, i.e. 56.18 %, were considered high and/or medium rated and was regarded as suitable for detecting inappropriate billing. With the help of a consultant, 162 negative rules were rated from high to low. The specialists who were discovered breaking these rules more, were often classified as high risk providers. These rule violations helped indicate how much one specialist deviated from their peers. Using a database called Program of Research Informing Stroke Management (PRISM), maintained by Medicare Australia and containing information from those specialists who were approached for previous compliance activities, were

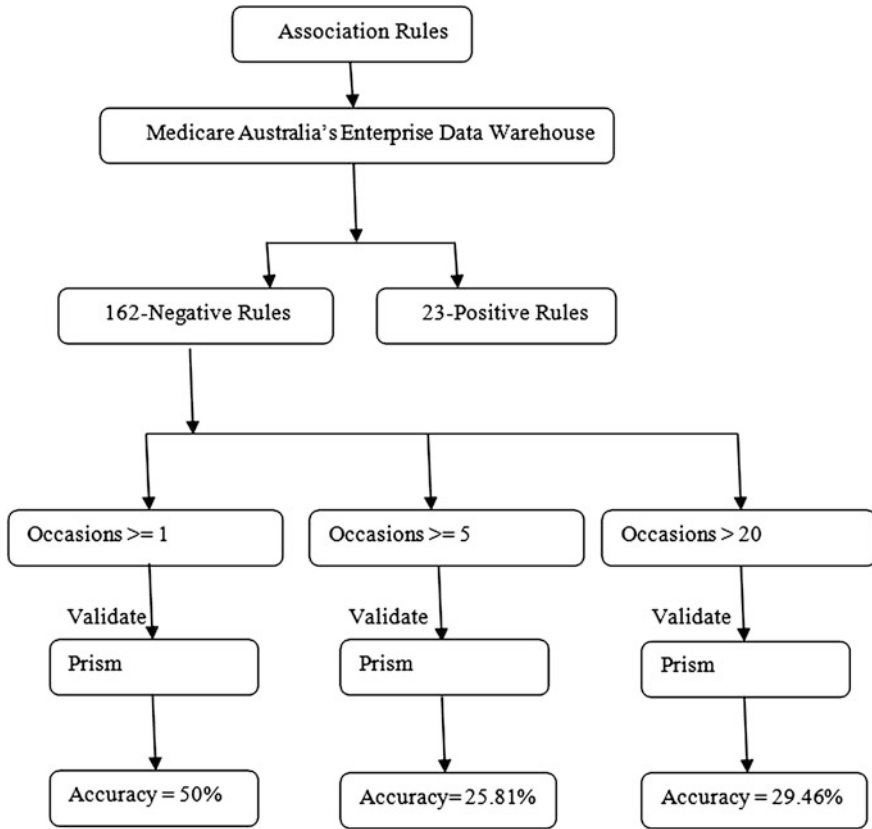


Fig. 12.10 Application of association rules

matched with suspected specialists who broke the rules. The specialists who broke rules were further divided into three classes:

1. Occasions ≥ 1 ,
2. Occasions ≥ 5 , and
3. Occasions > 20 .

It was found that ten specialists broke the rules on more than twenty occasions. Of these, five had a record in PRISM. This number showed the accuracy of these association rules to be 50%. For those who broke the rules on more than five occasions, the accuracy was found to be 25.81%, whereas the accuracy was 29.46% for the ones who broke rules on one or more occasion. Hence, these results suggest that even breaking one negative rule can be a good indication of noncompliant practice. Figure 12.10 demonstrated the application of Association Rules on PRISM database.

In another research conducted by Viveros et al. [14] association rules were applied to the episode database used for pathology services, in which the visit of each patient was linked to a record in the database. Hence, with a unique identifier, a database tuple was obtained. This tuple could consist of one or more medical tests performed in any instance of time with a maximum number of 20 tests per episode. Association rules were obtained with the setting of 50 % minimum confidence and values of 1, 0.5, and 0.25 % for minimum support. With a minimum confidence of 50 % and a minimum support of 1 %, 24 association rules were obtained. For a minimum confidence of 50 % and a minimum support of 0.5 %, 64 association rules were obtained. Further using a minimum confidence of 50 % and a minimum support of 0.25 %, 135 association rules were obtained. It was found that more knowledge of the behavior patterns was obtained by setting the minimum support 0.5 % rather than to 1 %.

12.4 Summary

In summary, the Supervised data-mining methods can be effective in discovering fraudulent transactions in healthcare system as they require accurate identification of fraudulent transactions and are trained to discriminate non-fraudulent and fraudulent transactions. Among all the supervised data-mining methods mentioned above, the neural network and association rules perform the best and, thus, are more often used by researchers to detect fraudulent patterns in healthcare data. Table 12.1 provides a summary of the supervised methods discussed in this chapter.

12.4.1 Advantages of Fraud Detection

The following are the advantages [25] that can be gained by efficiently detecting fraud in the healthcare system using information technology related tools like data mining and machine learning techniques:

- Yearly recovery of expenditures by the government and private sector for fraudulent claims,
- Reduction of the probability of fraudulent up-coding,
- Detection of new leads from increased accessibility of supplementary digital fingerprints,
- Authentication and confirmation of genuine services by call-back or web-based services,
- Digital authentication of services submitted by patients and providers,
- Decrease in record gathering time by use of a common identifier and growing digital media,

Table 12.1 A summary of supervised methods for fraud detection in healthcare

Algorithms based supervised method	Advantages	Disadvantages
Neural network	It requires less formal training. It has ability to sense complex non-linear relationships between independent and dependent variables. It also has an ability to detect all potential interactions between predictor variables, and the availability of multiple training algorithms [33]	It has black box type of nature and need high computation power. It can be prone to overfitting and its model development is of empirical nature [33]
Association rules	It can handle both table form data and transaction form data. It doesn't require the whole database to be fetch into the main memory [35]	Only database entries which exactly match the candidate patterns may contribute to the support of the candidate pattern. This creates a problem for databases containing small variations between otherwise similar patterns and for databases containing missing values [35]
Logistic regression	It is robust in nature and also independent variables does not required to be normally distributed, unbounded or interval. It can handle nonlinear effect [33]	It needs lots of data for achieving stable and meaningful results [33]
Classification tree	It has built in feature selection method. Models are easily interpretable, and are robust to the effect of outliers by easily dealing with missing values [34]	It can be unstable with small variation in the data. Some tree models can be very large and complex. It is expensive to be trained [34]
KNN classifier	As the entire process in KNN classifier is transparent, it is easy to implement and debug. It has special noise reduction technique which helps in getting accurate results [36]	For large training set it can have poor runtime performance as all the work in KNN classifier is done at run time. It is very sensitive to redundant features [36]
Genetic algorithm	It can rapidly scan a huge solution set. Bad proposals do not influence the end solution negatively as they are simply discarded. It works by its own rules without the need of knowing the specific rules of the problem. It is very useful for complex problems. Also its one of the main strength is the parallel nature of its stochastic search [38]	It is usually slower than traditional techniques [38]
Fuzzy logic Bayesian network	It allows the use of vague linguistic terms in the rules [37] It offers a natural and principled means of merging prior information with data. It also offers suitable settings for wide range of models [39]	It is difficult to estimate the membership function [37] It does specify about how it is selecting a prior and can generate posterior distributions that are greatly controlled by the priors. It is having high computational cost [39]

- Computerized digital verification for claims billing and payment,
- Real-time billing and confirmation of eligibility reimbursement,
- Decrease in customer time used in dealing with fraudulent claims,
- Prevention of duplication of imaging and laboratory tests,
- Prevention of creation of superfluous information which is already accessible digitally,
- Decrease in labor time to validate qualification,
- Decrease in material and effort to examine paper documentation,
- Decrease in time to store and retrieve paper records,
- Decrease in time used by a consumer in phone trees and taping redundant information,
- Decrease in communal medical fee and loss of life due to medication faults,
- Decrease in communal medical fee and loss of life due to clinical faults,
- Decrease in communal medical fee of duplicate diagnostic tests,
- Decrease in communal medical fee and loss of unnecessary medical surgeries,
- Decrease in referral visits to monitor prospective care provider through screening from a pay for performance initiative, and
- Decrease in provider time, bundling, accumulating, and forwarding of documentations to health plans, providers, and patients.

Hence, many benefits can be gained from new fraud detection techniques, which can help provide better medical services to authentic patients, help save money, and improve the healthcare experience for patients with genuine needs.

12.5 Conclusion

One of the most crucial problems facing the US government is fraud in healthcare system. Due to a large amount of data, it is impossible to manually audit for fraud. Hence, many statistical approaches have been proposed to overcome this problem. As fraud can be committed in complex and numerous ways, fraud detection is challenging, and there is a greater need for working models for fraud detection, including types of fraud that are not yet in use, as these models will not be outdated quickly.

To establish a well-functioning healthcare system, it is important to have a good fraud detection system that can fight fraud that already exists and fraud that may emerge in future. In this chapter, an attempt has been made to classify fraud in the healthcare system, identify data sources, characterize data, and explain the supervised machine-learning fraud detection models. Even though a large amount of research has been done in this area, more challenges need to be worked out. Fraud detection is not limited to finding fraudulent patterns, but to also providing faster approaches with less computational cost when applied to huge-sized datasets.

References

1. CMS (2011) Research, statistics, data and systems: national health expenditure data. NHE fact sheet
2. CMS (2011) Medicare: HCPCS—general information
3. FBI (2009) Reports and publications: 2009 financial crimes report
4. NHCAA (2007) The NHCAA fraud fighter's handbook: a guide to health care fraud investigations and SIU operations
5. IMF (2008) World economic and financial surveys: world economic outlook
6. Database NHCAA (2010) Combating health care fraud in a post-reform world: seven guiding principles for policymakers
7. NHCAA The problem of health care fraud, consumer alert: the impact of health care fraud on you, report of national health care anti-fraud association (NHCAA)
8. Koh H, Tan G (2005) Data mining applications in healthcare. *J Healthc Inf Mgmt* 19(2):64–72
9. OIG (2011) Medical fraud cases: OIG most wanted fugitive
10. He H, Hawkins S, Graco W, Yao X (2000) Application of genetic algorithms and k-nearest neighbor method in real world medical fraud detection problem. *J Adv Comput Intell* 4(2):130–137
11. Chan CL, Lan CH (2001) A data mining technique combining fuzzy sets theory and bayesian classifier—an application of auditing the health insurance fee. In: *Proceedings of the International conference on artificial intelligence*, pp 402–408
12. Ormerod T, Morley N, Ball L, Langley C, Spenser C (2003) Using ethnography to design a mass detection tool (MDT) for the early discovery of insurance fraud. In: *Proceedings of the ACM CHI conference*, 650–651
13. Ortega PA, Figueroa CJ, Ruz GA (2006) A medical claim fraud/abuse detection system based on data mining: a case study in chile. In: *Proceedings of international conference on data mining*, 224–231
14. Viveros MS, Nearhos JP, Rothman MJ (1996) Applying data mining techniques to a health insurance information system. In: *Proceedings of the 22nd VLDB conference*, Mumbai, India, pp 286–294
15. Yang WS, Hwang SY (2006) A process-mining framework for the detection of healthcare fraud and abuse. *Expert Syst Appl* 31:56–68
16. Liou F, Tang Y, Chen J (2008) Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Manage Sci*, 353–358
17. Shan Y, Jeacocke D, Murray D, Sutinen (2008) A mining medical specialist billing patterns for health service management. In: *Roddick J, Li J, Christen P, Kennedy P, (eds) Proceeding 7th Australasian data mining conference (AusDM 2008)*, Glenelg, South Australia. CRPIT, 87. ACS 105–110
18. Sokol L, Garcia B, West M, Rodriguez J, Johnson K (2001) Precursory steps to mining HCFA health care claims. In: *Proceedings of the 34th Hawaii International conference on system sciences*
19. Yang WS (2002) Process analyzer and its application on medical care. In: *Proceedings of 23rd International conference on information systems (ICIS02)*, Spain
20. Li J, Huang K, Jin J, Shi J (2008) A survey on statistical methods for health care fraud detection. *Health Care Manage Sci*, 275–287
21. Table F, Raineri A, Maturana S, Kaempffer A (2008) Fraud in the health systems of chile: a detection model. *Am J Public Health*, pp 56–61
22. Ghahramani Z (2004) Unsupervised learning
23. Rosella (2011) Predictive knowledge and data mining: healthcare fraud detection
24. Hall C (1996) Intelligent data mining at IBM: new products and applications. *Intell Softw Strateg* 7(5):1–11
25. Report on the use of health information technology to enhance and expand health care anti-fraud activities. Foundation of research and education of AHIMA

26. FBI (2011) Scams and Safety: common fraud schemes
27. London: The Guardian (2007) The mystery of John Darwin
28. Herb Denenberg (2005) The denenberg report: the insurance commissioners, other government agencies, and the insurance companies focus on insurance fraud committed by policyholders, but nothing is done about the multi-billion dollar racket of insurance fraud committed by insurance companies
29. Bhuvanewari R, Kalaiselvi K (2012) naive bayesian classification approach in healthcare applications. *Int j comput sci telecommun*, 3(1):106–112
30. Silver M, Sakata T, Su HC, Herman C, Dolins SB, O’Shea MJ (2001) Case study: how to apply data mining techniques in a healthcare dataware house. *J Healthcare Inf Manage* 15(2):155–164
31. Relles D, Ridgeway G, Carter G (2002) Data mining and the implementation of a prospective payment system for inpatient rehabilitation. *Health Serv Outcomes Res Method* 3(3–4):247–266
32. Anonymous (1999) Texas medicaid fraud and abuse detection system recovers \$2.2 million, wins national award. *Health Manag Technol* 20(10):8
33. Tu J (1995) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* pp 1225–1231
34. Lewis R (2000) An introduction to classification and regression tree (CART) analysis. Presented at annual meeting of the society for academic emergency medicine
35. Nayak J, Cook D (2001) Approximate association rule mining. In: Proceedings of the 14th International florida artificial intelligence research society conference
36. Cunningham P, Delany S (2007) k-Nearest neighbour classifiers. Technical report, UCD-CSI-2007-4
37. Russel S, Norvig P (2003) Artificial intelligence: a modern approach. Prentice-Hall, 2nd edition
38. Vose D (1995) The simple genetic algorithm: foundations and theory
39. Berger J (2006) The case for objective bayesian analysis. *Bayesian Anal* 1(3):385–402
40. Vick K (2009) As rescissions spawn outrage, health insurers cite fraud control. The Washington post, <http://www.washingtonpost.com/wp-dyn/content/article/2009/09/07/AR2009090702455.html>, Information Accessed on May 2012
41. Jeffries D, Zaidi I, Jong B, Holland M, Miles D (2008) Analysis of flow cytometry data using an automatic processing tool. *Cytometry Part A* 73A:857–867
42. Larose D (2005) Discovering knowledge in data, An introduction to data mining. Wiley InterScience
43. Niedermaye D (2008) An introduction to bayesian networks and their contemporary applications, innovations in bayesian networks. Springer, pp 117–130
44. De Jong KS, Spears WM, Gordon DF (1993) Using genetic algorithms for concept learning. *Mach learn* 13:161–188