International Journal of Hybrid Information Technology
Vol. 10, No. 8 (2017), pp.15-34
http//dx.doi.org/10.14257/ijhit.2017.10.8.02

Enriching Medicare Severity-Diagnosis Related Group (MS-DRG) Payments for Better Service to Inpatients using ANFIS

Kerina Blessmore Chimwayi, Noorie Haris,
Ronnie D. Caytiles* and N.Ch.S.N Iyengar**

School of Computer Science and Engineering, VIT University, Vellore, T.N., India
** Multimedia Engineering department, Hannam University, Daejeon, Korea*
***Professor, Department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, India*
kchimwayi@gmail.com, noorienittoor@gmail.com, rdcaytiles@gmail.com, srimannarayanach@sreenidhi.edu.in

Abstract

Variations in the cost for the same diagnosis among different hospital providers is a great concern to the public at large. With huge amounts of data being availed every second, utilising the data for the benefit of the society is commendable. In this research a neuro-fuzzy approach is proposed for Medicare payments data. Machine learning clustering algorithms on neuro-fuzzy results are compared to understand the variations in price for same treatment and diagnosis among different healthcare providers. Cluster analysis has been applied in various domains to help reveal hidden structures. Cluster analysis has not been well exploited in healthcare claims datasets, the reason being that healthcare expenditure data is highly skewed which make analysis complicated. The Inpatient charges is a large dataset that has 163065 and 12 attributes describing amounts paid by Centers for Medicare and Medicaid Services (CMS) to different healthcare providers using different Diagnostic Related Group (DRGs).

Keywords*: Clustering, Neuro-fuzzy, MS-DRGs, CMS*

1. Introduction

In this new era, healthcare costs are seen to be driving the demand for healthcare applications that are driven by big-data. In many countries healthcare spending outpaces Gross Domestic Product (GDP). The affordable Healthcare Act in the United States aims to improve healthcare through meaningful use of information technologies to reduce healthcare cost as well as to improve healthcare quality. The proposed model for this research is a great use case for all medical insurance provider companies and government medical insurance providers.

Currently big data is the greatest contributor for reducing the cost of fraud, waste and abuse in healthcare organizations as CMS. Large and unstructured datasets that characterize healthcare historical claims datasets are useful when machine learning algorithms are used for identifying useful patterns. Administration data in healthcare contains financial information around claims including other matters which is helpful when embedded with clinical data such as diagnosis codes and is overwhelmingly valuable for predictive models.

The purpose of this study is to identify cost change patterns for patients who are covered by Medicare so as to reveal hidden patterns about the providers as well as the DRG diagnosis represented. This kind of insight and analysis is of great use to the government and the stakeholders at large. Various clustering methods are implemented and compared at the end recommending the best clustering method for the highly dimensional non-numeric dataset. Clustering methods implemented include Partitioning

Copyright © 2017 SERSC

Around Medoids (PAM), Hierarchical clustering, K-modes and Self Organising Maps (SOM). Cluster analysis for the dataset will greatly aid in identifying the clusters of the most occurring DRGs among various states, identifying fraud as well as changes in costs amongst the providers. This research therefore employs ANFIS followed by clustering methods for CMS claims dataset so as to provide more transparency, accountability and affordability.

Clustering is a widely used technique for data reduction which is designed to uncover subgroups of observations within a dataset. Formally, clustering is defined as a technique of finding heterogeneous groups of data using a specific dissimilarity criteria. The result of successful clustering is a compact cluster whereby data items in one cluster are more similar to the members in the same cluster than members in the other clusters. Many researchers have worked on various clustering techniques in a bid to cluster medical data to find subtypes that lead to more targeted and effective treatments. Clustering techniques include hierarchical, partitioning and density based.

Meaningful clustering for the medical charges dataset helps to understand the characteristics of population, diseases and this offers extensive application prospects for various fields. Effective clustering in medical domain is of great help and is being explored to improve patient care and the overall quality service in the medical domain.

Majority of clustering algorithms require that mutually exclusive groups be identified. The constraint mentioned is difficult to satisfy in the case of real world medical datasets where mostly elements overlap. Applications of clustering in the medical domain include medical diagnosis and hospital resource management. Medical datasets to a greater extent carry mixed attributes, which therefore require appropriate clustering techniques since most clustering techniques consider numerical data.

In this research a neuro-fuzzy technique is implemented on inpatient charges and the results are clustered using various clustering approaches to compare and validate the results. This is done to understand in-depth the influences of various clustering algorithms on the medicare costs dataset, different methods are compared.

A combination of neural networks and fuzzy logic is generally defined as a system trained by a particular learning algorithm which is derived from neural network theory. Learning is done on the local information and modifications are done locally in the fuzzy system. This approach combine artificial neural networks (ANN) with Fuzzy Rule Based Systems (FRBS). FRBS is laid upon the structure of ANN and the learning algorithm is used to adapt FRBS parameters which can be membership functions. Adaptive Neural Fuzzy Inference System (ANFIS) method for implementing neuro-fuzzy on a medical charges dataset is considered.

Medicare Severity-Diagnosis Related Group (MS-DRG) is a method of putting a particular patient’s stay in a hospital in a group so as to facilitate payment for the services acquired. Clinical conditions that are similar in nature are put into one cluster. Each and every hospital discharge is given a certain relevant MS-DRG. Each provider identifier can have various records based on the number of these distinct clusters that are paid for. Inpatient Prospective Payment System (IPPS) per discharge payment considers national payment rates, operating expenses and capital expenses. Rates for payments are adjusted to take into consideration the costs tied to a particular patient’s health condition and DRG weight.

2. Literature Review

On a different perspective, authors in [1] proposed a method to tackle the issues related with medical data publishing. These problems include homogeneity attacks and the authors present a graded medical data publishing model which helps in using clustering to partition disease sensitive attribute values. The model helps improve security in the medical publishing of data. In [2] the authors highlighted on the use of medical imaging for census data.

Duong Thi Thu Huyen *et al*., [3] proposed a semi-supervised algorithm called as co-clustering technique for analysis of hospitals in terms of cost. Prior knowledge is required in the cases related to real world applications of dataset, so that they can be integrated in the clustering process. The most crucial factors for medical expenses were considered, those that reduced the cost gradually for any treatment at the same time improving the quality of services**.** With the help of these applications, decision-making is important in hospital management. Hospital-cost analysis among inpatients is an important aspect of yearly evaluation of cost in hospitals. New techniques having less computational complexity could be explored for clustering cost data of hospitals to derive more meaningful insights.

Richa Sharma *et al*., [4] identified patterns that were hidden and useful knowledge were extracted from database regarding medical details. For diagnosing diseases various classification and clustering algorithms have been used. Proposed model was implemented on the datasets of cancer diseases and heart diseases both of which are the complex diseases. Different approaches used in diagnosis of the diseases can be determined by the knowledge derived from the proposed model. Six powerful tools were used for the proposed model. Rapid Miner that is applicable for various business are highly efficient. Business applications, business analytics could also make use of framework. Manual coding to be done is minimal that served as an advantage to the users. R-Programming because of its expertise that is popular among the statisticians. It has the adaptability to implement various graphs. KNIME was used because of the popularity and portability of OS. It also allows plugin and extensions for meeting the requirements. It has also the capability to process large dataset.

Authors in [6] focus on comparing the results of what is termed statistical outliers with K-means for downsizing medical datasets. These researchers note the importance of proper pre-processing to accurate results in data mining. The results of the study proved that K-means performs better than Statistical outliers in reducing datasets in the presence of missing values.

Researchers in [7] have considered the k-means-mode for clustering medical data for the sake of predicting the likelihood of diseases. Authors distinguished the results for the K-mean-mode algorithm with background knowledge and without medical background knowledge. The results highlighted that medical knowledge nor hybrid clustering can perform well alone but effectiveness lies in combining both so as to produce excellent results.

Masumi Okuda *et al*.,[8] Surveyed patients at an emergency hospital in Japan were analyzed using distance for the exploration of the changes of the similarities observed in the factors of hospital regarding performance like, length of stay, hospital rebuilding and waiting time. Outpatient satisfaction structure consisted of those items that move around in hospital, those were personalized care and interpersonal skills. Inpatient satisfaction structure consisted of environmental items like the technical skills and interpersonal skills. When length of stay became longer, consolidation of groups took place to environment and the others. The revised survey after the completion of the survey of rebuilding of hospitals it was found that environmental items were not isolated and they were related to nurses and doctors that showed similarity. Distance usage for analysis has got a potential for exploring patient satisfaction hidden structure.

Sweta C. Morajkar *et al*., [9] proposed an approach of clustering of temporal data using evolutionary clustering. Time dimension was however not considered. Normally it is seen that clustering techniques usually focus on grouping data objects on the basis of similarity function. Temporal data clustering is an extension of traditional clustering techniques and determines solutions for discovering the information over a time period. A methodology for clustering patients' medical data as proposed based on a new similarity measure. They have avoided unnecessary distance calculations thereby accelerating the clustering technique on application of such similarity measure.

In their study [10] considered blended clustering in healthcare data mining. The research observed that in terms of hospital utilization in Australia, length of stay, legal status, age as well as economic situation has an effect in usage of health services. In this approach k-means has been used for a patient profile dataset whereby various patterns of utilizing hospital service for a different population having a wide range of culture beliefs and age were considered.

B.Simhachalam *et al.*, [11], proposed a model that classified the patients suffering from thyroid, on the basis of clustering approach called as Possibilistic Fuzzy C-means clustering. Results that were obtained were compared with Fuzzy c-Means clustering algorithm according to the performance of classification. From the comparisons made, it showed that the Possibilistic Fuzzy C-means clustering technique performed well. The limitations of these clustering methods include issues of running several times to obtain better classification results.

In a certain study [12] researchers proposed mining sensor data so as to control inpatient environment. Simple K-means has been compared to Hierarchical clustering. Results from the study proved that hierarchical clustering works better than simple K-means. Hierarchical clustering here is used to classify each patient in certain cluster whereby the level desired for temperature and light values at a specified time period is achieved. This method performs well because of the ability to learn incrementally.

Fast search and find of density peaks (FSFDP) was suggested and implemented as noted in [13]. The notion of lower or higher local density is greatly used to describe clusters. Most researchers have considered clustering in the medical fraternity using only the University of California Irvine (UCI) benchmark and real life datasets available from UCI. Researchers proposed that the clustering methodology is suitable for real world practical applications. Dermatology dataset has been exploited using MATLAB.

In [14] researchers proposed the use of logistic regression for predicting and identifying the acute care patients who might be at risk of being re-admitted. The dataset used in this study is a healthcare claims dataset. Adjusted Clinical Group (ACG) model has been used for generating features focused on an individual for a certain inpatient and outpatient period of time. Aggregated Diagnostics Groups were also considered since these greatly assess an individual’s health through the use of prescribed drugs and diagnosis.

Christina Klüver , [15] introduced a neural network called as Self Enforcing neural network (SEN) that learns through self-organization , to cluster data containing medical details. A validity factor called cue was also considered that affected the clustering of the data. On analysis of results, it was observed that the user can be influenced by the clustering of data through SEN, in turn making data analysis to be dependent on economical and medical interests. Prototype proposed includes concrete examples and shows network’s potential for analysis of complex medical data. Greatest limitation is that SEN needs to be reprogrammed according to the latest development to process complex medical data.

Sankalp Khanna *et al.*, [16] describes an approach of applying a clustering technique on the basis of time on data regarding health details for creating visualizations and patient flow analysis. Clustering of inpatients resulted into slots on hourly basis, and were grouped on the basis of certain parameters that resulted into a tool that can visualize and analyse interdependencies and interactions between parameters of hospital patient flow. They could drive simulations for analysis of strategies like discharges that happen at early stages and implement strategies for management of hospital services individually.

In their research [17] proposed a hybrid model K Harmonic Means and Overlapping K-Means (KHM-OKM) for clustering medical data whereby the output of KHM is used as input to initialize the cluster centers for OKM. These researchers identified that medical datasets usually have overlapping information which required an improved overlapping k means algorithm for handling the unique nature. The results in this study proved that the hybrid method is better than OKM alone for clustering 10 medical datasets available from UCI.

Researchers in [18] explored the use of data mining in healthcare informatics. Authors carried an extensive literature survey and identified that in healthcare machine learning techniques are much more used as compared to data mining from the period of 2004 to 2015. These various applications have gone a long way in improving healthcare costs, service, treatments and quality decision making.

Ali K. Hmood *et al.*, [19] studied the problem regarding medical reports to be anonymised and thereby produce a solution to preserve the information utility while at the same time anonymising a collection of the medical reports. These were done for the purpose of cluster analysis of medical reports. Experimental study on real time medical reports suggested that proposed model can maintain the information utility and privacy protection in medical reports. Data mining for high quality results require access to patient record information. Releasing patients' medical reports may lead to revealing of sensitive information of individual patients.

Authors have focused on clustering of medical reports. Clustering techniques like PAM, C-means and K-means were applied on medical reports before and after anonymization of the data. The study was not focused on the clustering techniques but analysis of clusters from text documents of the tasks that were performed on medical reports that were published after anonymization. The disadvantage of the study was its inability to guide the anonymization process as there were no class labels in clusters.

Pavlos Delias *et al.*, [20] supported decision-making for flexible environments with help of logical process models. A mining technique was proposed as a technique to cluster customer flow and thereby create effective summarization. A similarity metric was created for efficient downgrading of effect of noise. A spectral technique emphasizing the estimated groups' robustness was used, that provided clearer maps of the process involved. Proposed method was applied on a real time healthcare institution that delivered valuable results and showed better performance of process models' density and complexity. But there are environments that exists where in there is flexibility in customer flow and such models are hard to be build.

Authors in [21] predicted future admission orders based on historical subsets of training data. It has been noted that giving a higher priority to small amounts of data which are recent is better than using larger amounts of older data when dealing with future clinical predictions. In this study the main objective was to determine the varying longitudinal training data and the impact it has on the future of clinical decisions.

To address the challenge of high dimensionality and inherent sparseness in clustering medical data, [22] opted for the use of Multiple Level Clustering whereby meaningful pre-processing is performed then the data is partitioned to form cohesive groups which will be analyzed locally. The researchers also proposed mobile application for patients which enables data objects to be clustered to a same group whilst acquiring domain expert feedback for the suggested classification.

Moumita Bhattacharya *et al.*, [23], discussed the factors that affect obesity and detection of likelihood of obesity. Children having higher risk of obesity were identified through identification of unique patterns in them. The methods that they have used are clustering methods that groups children together whose body measurements are alike within a period of time. Measurements observed has grouped children belonging to the same cluster that were then plotted against age. Different clusters formed are associated and are used to separate children into the groups of higher, middle, or lower cluster based on measurements of growth pattern s.

K-means and EM based clustering approaches were used as their methods to proposed work to obtain distinct groups based on different measurements. Children belonging to the higher category of growth pattern were tagged with weight that represents it as highest and were at risk of obesity. Comparison of growth patterns were made to demonstrate the utility of their approaches. They could have even established these differences more clearly by considering even the non-temporal attributes.

In all the research work reviewed, a gap exist in that many clustering methods were done with respect to medical datasets, however none has explored the used of machine learning and soft computing techniques to cluster and analyze Medicare charges data to see the variability in diseases that are in the same DRG.

3. Methodology

3.1. Problem Identification

Variability in healthcare charges for the same disease is a great concern to the public at large who deserve fairness and high quality in healthcare services.

3.2. Data Collection

The medical charges dataset under consideration is available from data.gov. The data has information pertaining payments made by Medicare to various medical providers for the top 100 DRGs in all states. The dataset is of high dimensions and has categorical and continuous attributes. The dataset has 163,065 instances and 12 attributes.

Table 1. Dataset Variables

Attribute	**Description**
Definition of DRG	MS-DRG can be identified by the description and code.
Id of provider	The Medicare certified hospital facility has got CMS Certification Number assigned to it.
Name of provider	Provider name for each of the DRGs.
Street address of provider	Street address of provider.
City of Provider	Location of the provider in terms of city.
State of provider	Location of the provider in terms of states.
Zip code of provider	Zip code of the provider
HRR of provider	Location of provider in terms of Hospital Referral Region.
Total Discharges	This refers to the sum of all discharges made by provider
Average Covered Charges	Medicare services charged on average by the provider for all discharges belonging to DRG.

Copyright ⓒ 2017 SERSC

Average Total Payments	The payments on the basis of average total to all providers for the DRG including the DRG capital, outlier payments, amount, disproportionate share and teaching for all cases.
Average Medicare Payments	This is the average amount that Medicare pays to the provider and the provider gets paid on average basis by the Medicare for share of Medicare of the DRG.

The Figure.1 below describes the flowchart for the proposed methodology. Structure of healthcare dataset before pre-processing has been extracted and is shown below in fig 2. The structure is a combination of factors, integers and numbers.

3.3. Pre-processing of Data

To discover meaningful knowledge the medical data must be transformed into the most appropriate format for analysis. The medical charges dataset under consideration is highly dimensional and has mixed attributes which makes it difficult for various machine learning techniques to be applied. Dataset requires pre-processing as it contains a mixture of attribute data types. Attributes like the DRGs and provider states need to be encoded into numerical data that makes it suitable for the application to the developed model. Since there are cost variations in the context of average covered charges for the same DRG groups belonging to different states, addition of a class explaining the risk of being paid more to the hospitals for the inpatient discharges, makes the dataset even more meaningful. An outlier threshold average covered charges was set as $24,758 as per the fiscal year 2014. Beyond this threshold payment, it is considered to be a high risk possibility of fraud affected in the case of the corresponding provider of the associated DRG group who provides Medicare services for the inpatients. If average covered charges is greater than the threshold payment, then the presence of risk is pre-coded as 2 else 1.

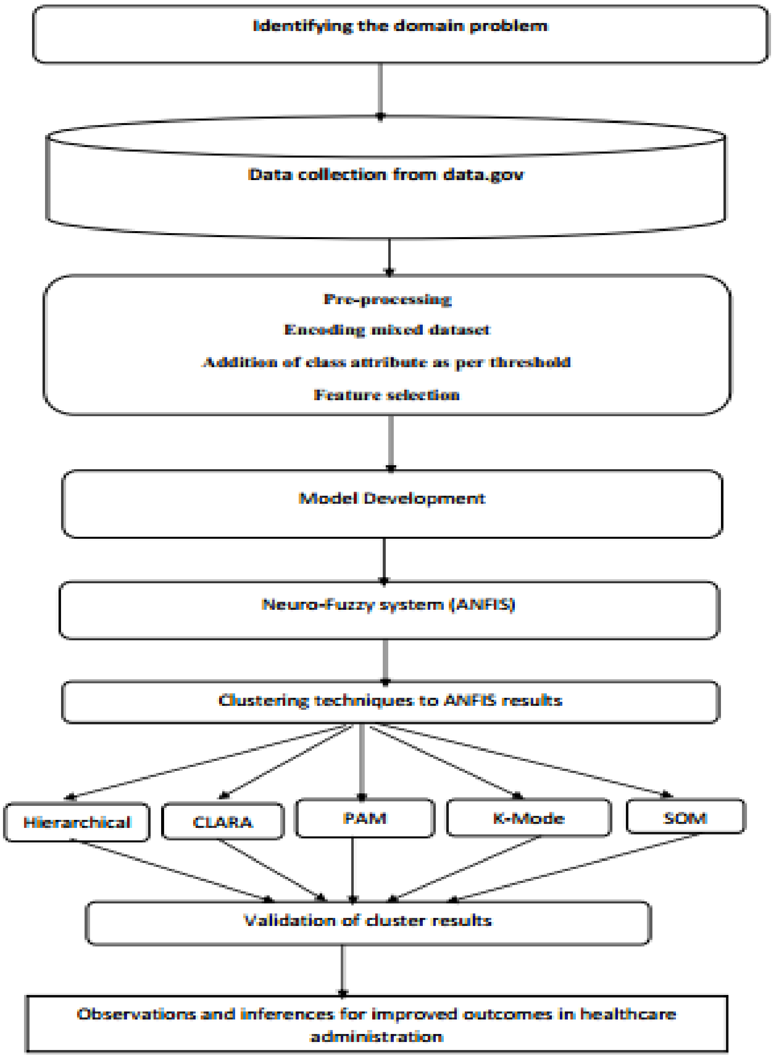
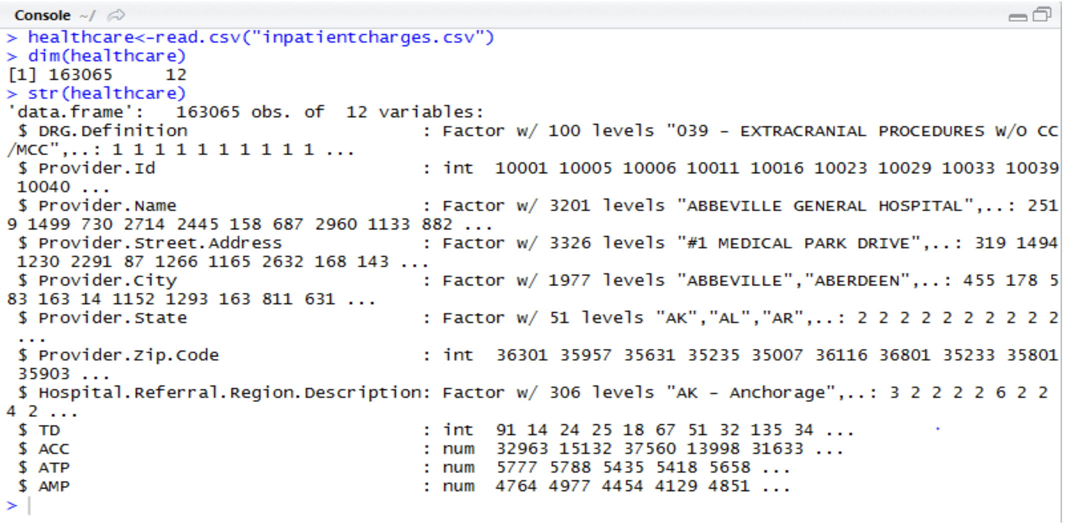

Identifying the domain problem
Data collection from data.gov
Pre-processing
Encoding mixed dataset
Addition of class attribute as per threshold
Feature selection
Model Development
Neuro-Fuzzy system (ANFIS)
Clustering techniques to ANFIS results
Hierarchical
CLARA
PAM
K-Mode
SOM
Validation of cluster results
Observations and inferences for improved outcomes in healthcare administration

Figure 1. Methodology Flowchart



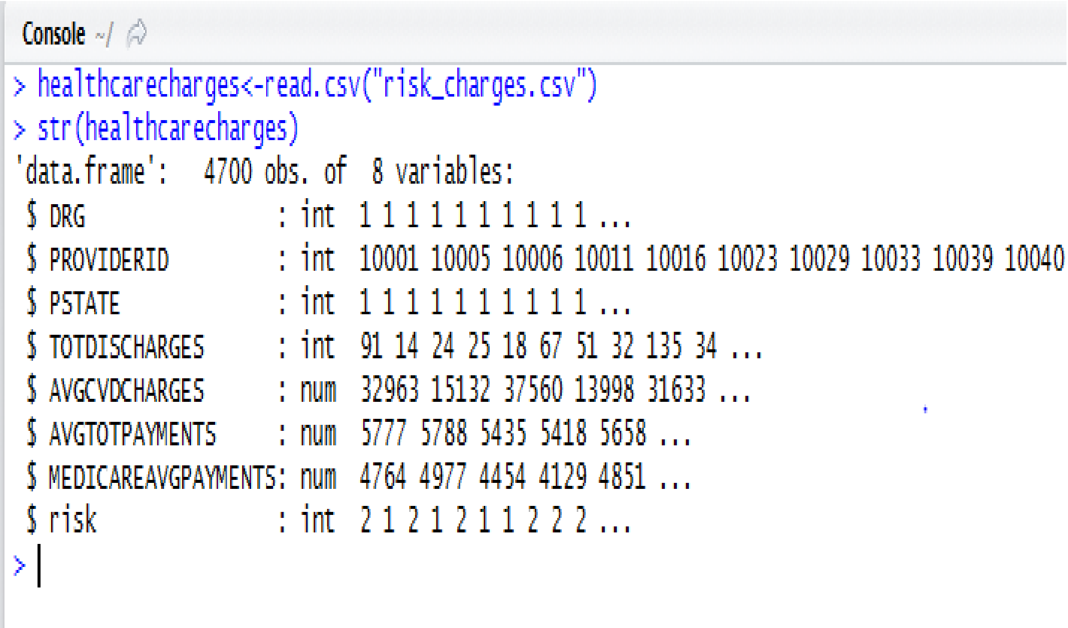
```
> healthcare<-read.csv("inpatientcharges.csv")
> dim(healthcare)
[1] 163065     12
> str(healthcare)
'data.frame':   163065 obs. of  12 variables:
 $ DRG.Definition                      : Factor w/ 100 levels "039 - EXTRACRANIAL PROCEDURES W/O CC
/MCC",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Provider.Id                         : int  10001 10005 10006 10011 10016 10023 10029 10033 10039
 10040 ...
 $ Provider.Name                       : Factor w/ 3201 levels "ABBEVILLE GENERAL HOSPITAL",..: 251
9 1499 730 2714 2445 158 687 2960 1133 882 ...
 $ Provider.Street.Address             : Factor w/ 3326 levels "#1 MEDICAL PARK DRIVE",..: 319 1494
 1230 2291 87 1266 1165 2632 168 143 ...
 $ Provider.City                       : Factor w/ 1977 levels "ABBEVILLE","ABERDEEN",..: 455 178 5
83 163 14 1152 1293 163 811 631 ...
 $ Provider.State                      : Factor w/ 51 levels "AK","AL","AR",..: 2 2 2 2 2 2 2 2 2 2
 ...
 $ Provider.Zip.Code                   : int  36301 35957 35631 35235 35007 36116 36801 35233 35801
 35903 ...
 $ Hospital.Referral.Region.Description: Factor w/ 306 levels "AK - Anchorage",..: 3 2 2 2 2 6 2 2
4 2 ...
 $ TD                                  : int  91 14 24 25 18 67 51 32 135 34 ...
 $ ACC                                 : num  32963 15132 37560 13998 31633 ...
 $ ATP                                 : num  5777 5788 5435 5418 5658 ...
 $ AMP                                 : num  4764 4977 4454 4129 4851 ...
> 
```

Figure 2. Structure of the Data before Pre-Processing

Copyright ⓒ 2017 SERSC

The Dataset obtained an additional attribute of risk as a class attribute where each row cloud be classified as high risk (2) or low risk (1). Removal of the attributes namely provider name, provider Street Address are due to the reasons that they convey the same details as conveyed in terms of provider id and provider state. Therefore the essential attributes for the input to the neuro-fuzzy system are DRG, Provider ID, Provider state, total number of discharges, the average charges covered, average of the total charges and Medicare payments received.

The new pre-processed dataset with the structure is presented in Figure 3. All the attributes after pre-processing have been transformed to integers and numbers.



```
> healthcarecharges<-read.csv("risk_charges.csv")
> str(healthcarecharges)
'data.frame':   4700 obs. of  8 variables:
 $ DRG               : int  1 1 1 1 1 1 1 1 1 1 ...
 $ PROVIDERID        : int  10001 10005 10006 10011 10016 10023 10029 10033 10039 10040
 $ PSTATE            : int  1 1 1 1 1 1 1 1 1 1 ...
 $ TOTDISCHARGES     : int  91 14 24 25 18 67 51 32 135 34 ...
 $ AVGCVDCHARGES     : num  32963 15132 37560 13998 31633 ...
 $ AVGTOTPAYMENTS    : num  5777 5788 5435 5418 5658 ...
 $ MEDICAREAVGPAYMENTS: num  4764 4977 4454 4129 4851 ...
 $ risk              : int  2 1 2 1 2 1 1 2 2 2 ...
> |
```

Figure 3. Structure after Encoding

3.4. System Design

3.4.1. Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

ANFIS works similarly to fuzzy inference system. Architecture of ANFIS for the inpatient discharges payments can be depicted as shown in the Figure 4. There are five linguistic labels of membership function for each of the attribute given as input to the neuro-fuzzy system name, very small, small, medium, large and very large. All nodes in layer 1 are the adaptive nodes. Membership function can take the shape of gaussian, trapezoidal or triangle. For the proposed model gaussian function has been used. Gaussian function is represented by two parameters namely membership value and the membership function range.

The Gaussian membership function for all the attributes after passing on to the layer 1 are as shown in Figure 5. For all the nodes which are in layer 3, the node represent the output which is a product obtained from incoming signals in the second layer. The fourth layer has a fixed label termed. Particularly every node in this layer calculates the ratio for the node's rule's firing strength totalling to the sum of every rules' firing strengths. All nodes in fourth layer adaptive nodes which have a node function. In the fifth layer a single node is fixed and labelled as risk of the variation in inpatients' discharge payments, which is computed by the summation of all incoming signals from the layer 4. The output thus obtained is the defuzzified output.

Copyright ⓒ 2017 SERSC

The rules below describe the working of the algorithm adopted.

Rule 1:

If a is P_1 and b is Q_1 then g1=x_1a+y_1 b+z_1.

Rule 2:

If a is P_2 and b is Q_2 then g1=x_2a+y_2 b+z_2.

The layers of this ANFIS each work differently to reach the targeted and expected outcome. The working is therefore explained below.

Layer 1:

- $Z_{L,j}$ resembles the output from jth node of the layer L.
- The layers for this are adaptive nodes which have a node function associated to it
  $Z_{L,j} = \mu P_i$ (a) for j = 1, 2, or
  $Z_{L,j} = \mu Q_{i-2}$ (a) for j = 3, 4
- a (or b) are the input nodes j and Pj (or Qj−2) resembles the linguistic labels which are associated with this node.
- $Z_{L,j}$ is the membership function of fuzzy set (P1, P2, Q1, Q2).

Membership function µ (a)= $\frac{1}{1+\left|\frac{a-r_j}{p_j}\right|^{2q_j}}$ (1)

P_j,q_j and r_i are premise parameters.

Layer 2:

- All nodes in this layer are fixed nodes labelled p
- Product of all the incoming signals are forwarded to the output.
  $Z_{2,j} = W_j= \mu P_j$ (a). μQ_j (b) for j = 1, 2
- Every node in this layer represents the firing strength of the rule.
- T-norm operator is used like AND operator.

Layer 3:

- All nodes in this layer are fixed nodes that are labelled as Norm.
- These node calculate the ratio of the j^{th} rule's firing strength to the sum of all rules' firing strengths.

$Z_{3,j}= \overline{W_j} = \frac{W_j}{W_1+W_2}$, j = 1, 2 (2)

- Normalized firing strengths represents the output.

Layer 4:

- All nodes belonging to this layer are adaptive nodes associated with a node function:

$Z_{4,j} = W_j\ G_i = W_j(x_ja+y_j\ b+z_j)$ (3)

- Layer 3 produces the normalized firing strength, W_j .
- The parameter set of node is {x_j , y_j , z_j } in this case are referred to as consequent parameters.

Layer 5:

- This layer contains a single node which is a fixed node labelled as S. All the incoming signals are given to output as the summation of all signals
  - Output of layer 5 = $Z_{5,j} = \sum_j \overline{W}_j g_j = \frac{\sum_j W_j g_j}{\sum_j W_j}$ (4)

The architecture below shows the working of ANFIS for the CMS dataset under consideration. The inputs are the attributes taken after feature selection which include the DRG among others which are therefore processed to layer one which contain linguistic variables derived from the model. Layers 3 and 4 have signals and various activations whereas layer 5 is the output node.

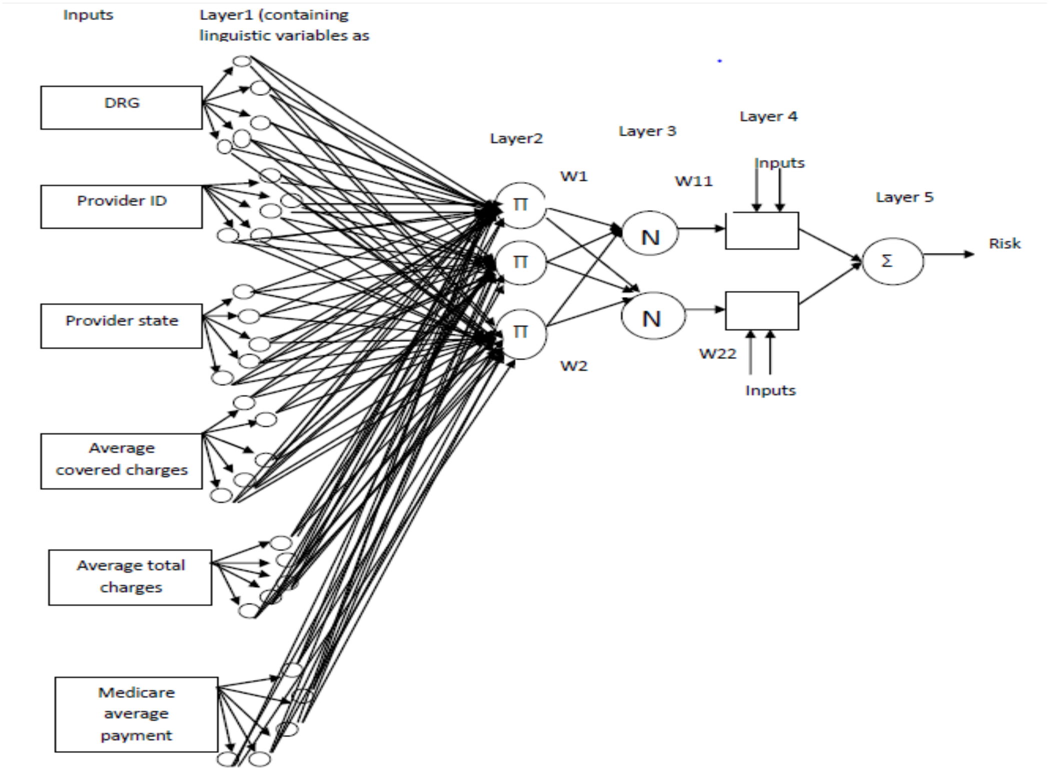


Figure 4. Neuro-fuzzy Architecture

The diagram below gives an overview of the membership functions derived from ANFIS model. The membership functions show the Gaussian membership for each attribute in the model. Rules from these membership functions are formed with 5 linguistic variables namely very small, small, medium, large and very large.

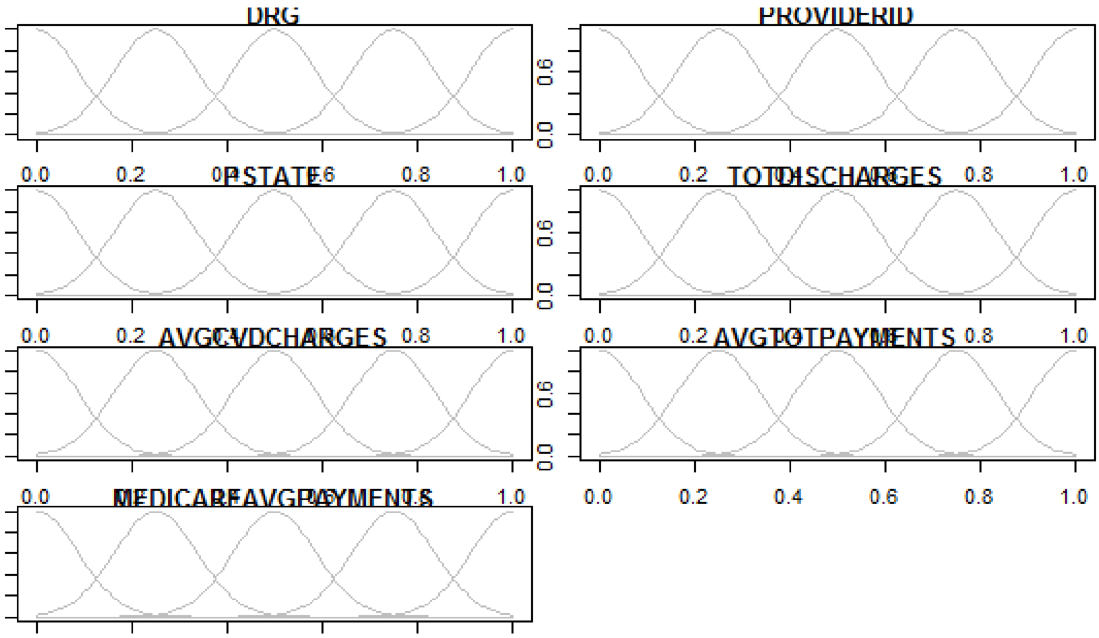


Figure 5. Gaussian Membership Functions

3.4.2. Clustering Techniques

Various clustering techniques have been applied on the dataset. Number of clusters required for the cluster is determined by the elbow method of k-means clustering that gave a result of 3 as the optimal number of clusters as inferred from the graph of Figure 6.

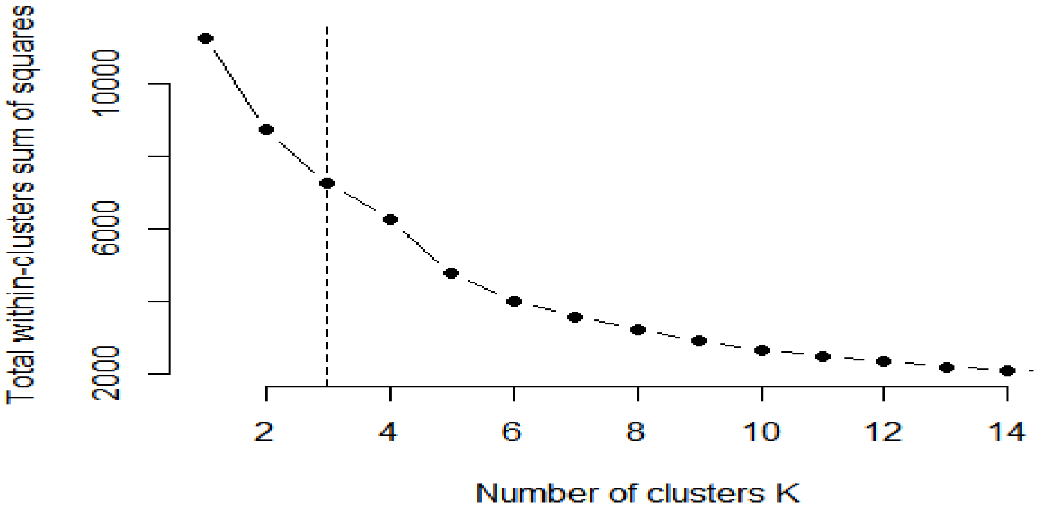
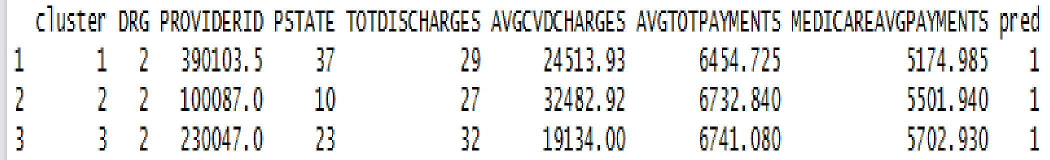


Figure 6. Elbow Method for Determination of Number of Clusters

After determination of the number of clusters various clustering techniques are applied on the dataset.

3.4.2.1. Hierarchical Clustering

On application of hierarchical clustering technique on to the dataset gave the classification results of risks involved in patients as shown in Figure 7. The clusters range from 1 to 3 and using the features in the medical dataset which have been applied to ANFIS are clustered and the figure below describes in detail the members in each particular cluster. Of particular interest is the predicted class as for these clustered elements.



```
  cluster DRG PROVIDERID PSTATE TOTDISCHARGES AVGCVDCHARGES AVGTOTPAYMENTS MEDICAREAVGPAYMENTS pred
1       1   2   390103.5     37            29      24513.93       6454.725            5174.985    1
2       2   2   100087.0     10            27      32482.92       6732.840            5501.940    1
3       3   2   230047.0     23            32      19134.00       6741.080            5702.930    1
```

Figure 7. Hierarchical Clustering Results

3.4.2.2. K-modes Clustering

K modes clustering works on mixed attributes. It is based on the statistical calculation of mode of every attributes in the dataset *i.e.*, the frequently occurring type of attribute will be made as the centroid of each of the clusters that are formed. On application of k modes clustering on the dataset, it gave the classification results as shown in the figure.

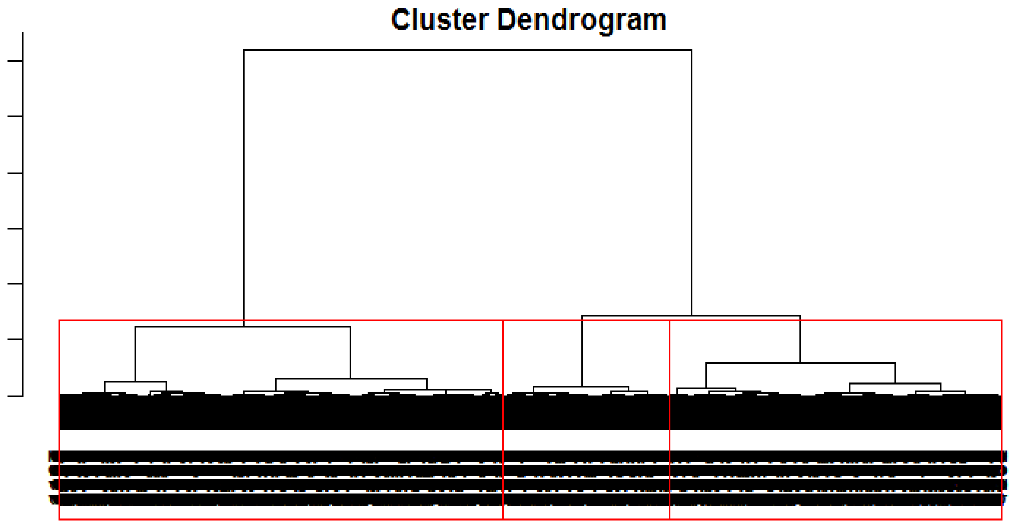
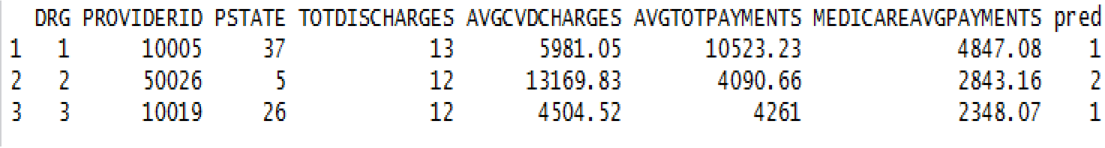


Figure 8. Dendogram for Classified Results



```
  DRG PROVIDERID PSTATE TOTDISCHARGES AVGCVDCHARGES AVGTOTPAYMENTS MEDICAREAVGPAYMENTS pred
1   1      10005     37            13       5981.05       10523.23             4847.08    1
2   2      50026      5            12      13169.83        4090.66             2843.16    2
3   3      10019     26            12       4504.52           4261             2348.07    1
```

Figure 9. K-Mode Clustering Results

3.4.2.3. SOM Clustering

SOM clustering is an unsupervised clustering technique wherein it takes whole of the dataset for training the network. All the instances of the healthcare dataset are considered as the node and then the similarity between the nodes is learnt by the network at the time of training.

SOM clustering when applied on the dataset gave the classification of the dataset based on the similarity among the instances of the dataset as in the Figure 10. The colors resemble different clusters formed which are 3 represented by orange green and blue.

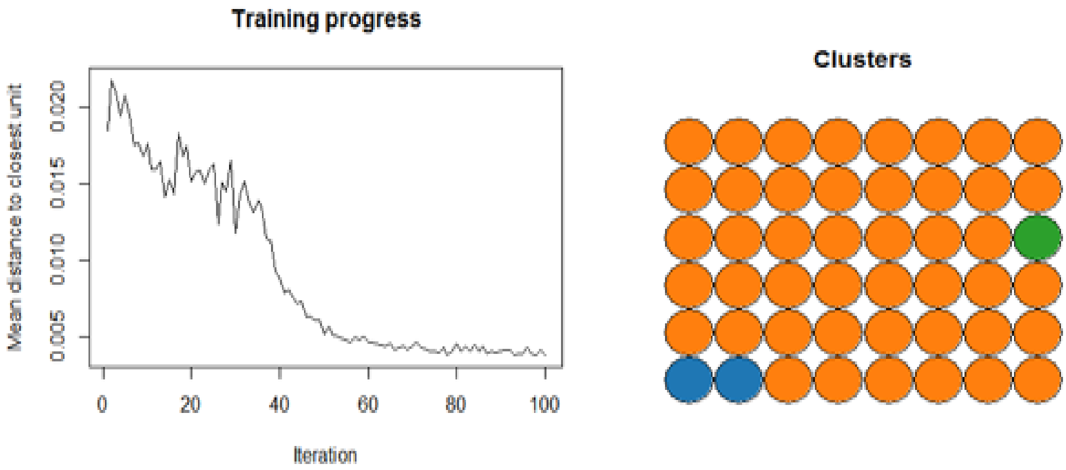
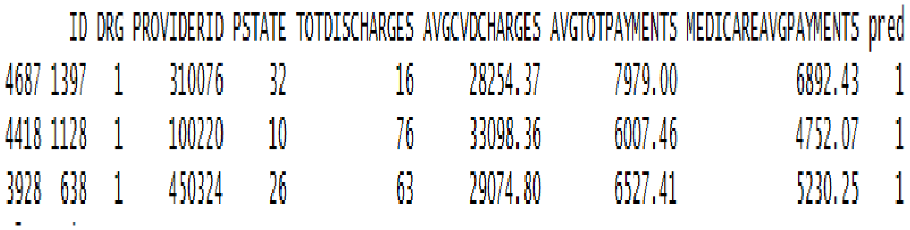


Figure 10. SOM Training and Clusters

3.4.2.4. PAM Clustering

PAM clustering finds a sequence of instances of the dataset called medoids that are located as the centers in clusters. The aim of the algorithm is minimizing the average dissimilarity of instances to their closest selected instance. Thus, minimizing the sum of dissimilarities between an instance and its closest selected instance. PAM clustering result in the classification of risks as shown in the Figure 11and 12.



```
       ID DRG PROVIDERID PSTATE TOTDISCHARGES AVGCVDCHARGES AVGTOTPAYMENTS MEDICAREAVGPAYMENTS pred
4687 1397   1     310076     32            16      28254.37        7979.00             6892.43    1
4418 1128   1     100220     10            76      33098.36        6007.46             4752.07    1
3928  638   1     450324     26            63      29074.80        6527.41             5230.25    1
```

Figure 11. PAM Clustering Results

PAM clustering plot shown below gives a visualization of the clusters for the Medicare inpatients. It is clear from the figure that clusters are overlapping. This is because of the nature of medical data sets whereby most features overlap. However, we can identify 3 clusters represented.

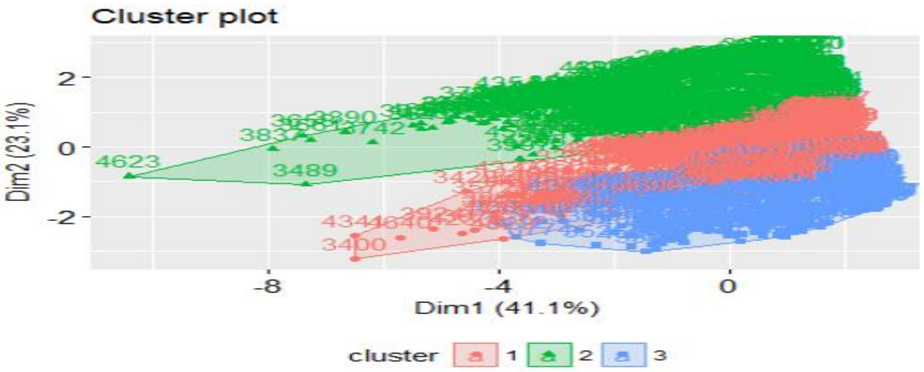
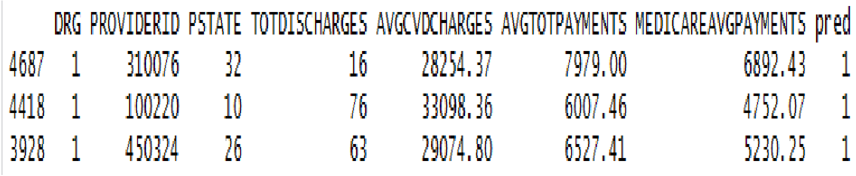


Figure 12. PAM Clustering Plots

3.4.2.5. Clustering Large Applications (CLARA)

CLARA clustering technique works on larger datasets. It subsets the dataset and each of the subset. Each sub-dataset creates the medoids which are obtained from the most useful sub dataset. To reach the appropriate sample size, each random observation is added to the set. CLARA clustering resulted in the classification of the risks as medoids as shown in Figure 13.



```
     DRG PROVIDERID PSTATE TOTDISCHARGES AVGCVDCHARGES AVGTOTPAYMENTS MEDICAREAVGPAYMENTS pred
4687   1     310076     32            16      28254.37        7979.00             6892.43    1
4418   1     100220     10            76      33098.36        6007.46             4752.07    1
3928   1     450324     26            63      29074.80        6527.41             5230.25    1
```

Figure 13. Clara Clustering Results

The plot for CLARA shows clear demarcations between the 3 clusters formed.

Copyright ⓒ 2017 SERSC

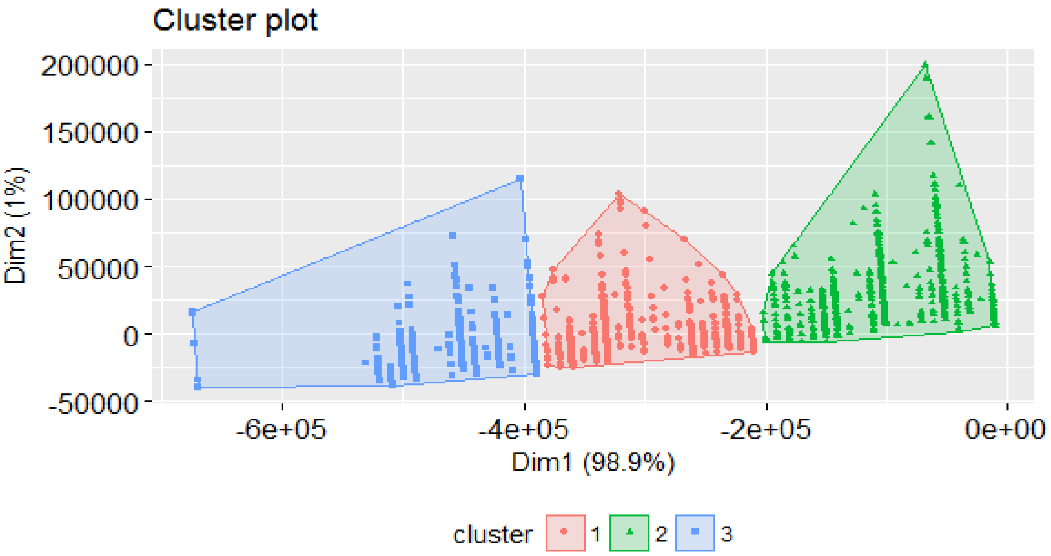
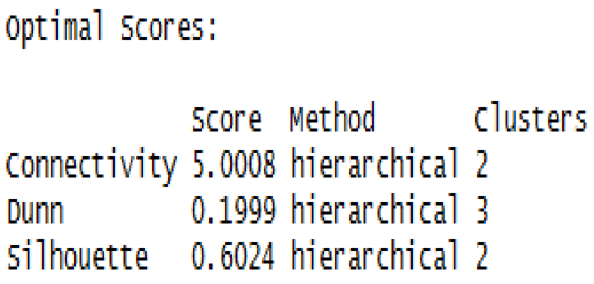


Figure 14. CLARA Clustering Plot

4. Validation

4.1 Cluster Validation

Validation of clustering is to identify the suitable clustering technique to be applied for the data. It is found that hierarchical clustering is found to be suitable for the dataset because of the parameters that explain the reason of it being suitable. The parameters are connectivity, dunn and silhouette. Results can be summarized as in the Figure 16.



```
Optimal Scores:

             Score  Method       Clusters
Connectivity 5.0008 hierarchical 2
Dunn         0.1999 hierarchical 3
Silhouette   0.6024 hierarchical 2
```

Figure 15. Cluster Validation Results

Connectivity shows the extent to which data instances will be placed in a same cluster with respect to the nearest neighbors. The value is allowed to be from 0 to infinity. Cluster validation also considers Dunn index which is defined as a ratio between smallest distances among instances of data in different cluster in comparison to the largest intra-cluster distance. It has a value between 0 and infinity which should must be maximized. Average Silhouette width lies between -1 (in case of poorly clustered instances of data) to 1 (well clustered instances of data) which should be maximized.

5. Results and Discussions

The confusion matrix for the neuro-fuzzy algorithm when it was applied on the dataset gave the following results of classifying the risk as indicated in Table .2.

Copyright ⓒ 2017 SERSC

Table 2. Confusion Matrix

Actual	Predicted	
	1	2
1	660	5
2	236	509

The error rate for prediction is found to be 17.0922. Accuracy of the neuro-fuzzy model can be summarized as in Table 3. A higher accuracy proves that the model performs well on the dataset.

Table 3. Accuracy Measures for ANFIS

Accuracy	83%
Sensitivity	100%
Specificity	97%

The plot shows the accuracy measures for the model. The model has high sensitivity and specifity.

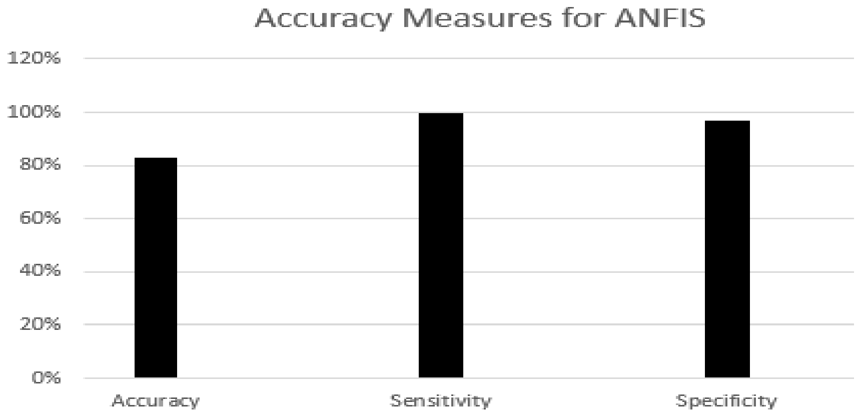


Figure 16. Plot for ANFIS Accuracy

Upon cluster validation, hierarchical clustering has been chosen as the best method for the dataset. Table 4 below therefore summarizes the 3 clusters that were formed, all falling into one DRG.

Table 4. Cluster Description

Cluster	Analysis
Cluster 1	For DRG 2 Intracranial Hemorrhage or Cerebral Infarction W MCC, the state of Ohio represented by 37 has a total discharge of 29, average covered charges of 24 513 have a lower risk represented by 1.
Cluster 2	In this cluster, the state is Florida and the average covered charges are greater than the threshold but because of the lower Medicare payments received it has been clustered into the DRG 2 and classified as low risk
Cluster 3	For DRG 2 Intracranial Hemorrhage or Cerebral Infarction W MCC, the state of Michigan represented by 23 has a total discharge of 32, average covered charges of 19 134 which is lower than the threshold have a lower risk represented by 1

Copyright ⓒ 2017 SERSC

6. Conclusions and Future Work

In this paper, ANFIS was implemented and various clustering techniques were explored on ANFIS results so as to identify unique and important patterns in Medicare payments data. Experiment results show that for the data, hierarchical clustering is the best. Effective clustering for administrative medical datasets, in this case the Medicare payments is crucial to the development of the healthcare domain since it goes a longer way in identifying useful patterns for assessing quality, costs, fraud and to a greater extend variability in the costs of DRGs. The dataset under consideration is a great limitation since useful attributes for prediction were unavailable. The size of the dataset is huge to be handled by mere computers but requires big data tools as Hadoop and Mapreduce or Scala. Future work therefore requires more meaningful prediction of yearly payments only if the dataset is added with features that further explain variability is costs. The work served as a basis for estimating inpatient charges for healthcare insurance providers. Big data applications serve as a necessity for the improvement and better application of the work considered.

Acknowledgements

The author Dr.N.Ch.S.N.Iyengar is thankful to Dr.P.Narasimha Reddy, Executive Director, SNIST and Dr.G.Viswanathan, Chancellor, VITU for their support and encouragement.

.

References

[1] T. Yi., M. Shi, W. Shang and J. Cao, Graded Medical Data Publishing Based on Clustering, **(2015)**, pp. 1647–1652.

[2] B. Yahaya , R. Latip, A. Abdullah and M. Othman, Medical Data Simulation with Self-Adaptive Multi-Instance Broker in Hierarchical Cluster Grid Structure. **(2015)**. Accessed online from https://doi.org/10.1109/CASH.2014.24.

[3] D. Thi, T. Huyen, L.H. Son and A. Drogoul, Semi-supervised fuzzy co-clustering for hospital-cost analysis from electronic medical records, **(2016)**, pp. 25–30.

[4] R. Sharma, Medical Data Mining Using Different Classification and Clustering Techniques: A Critical Survey. Accessed online from https://doi.org/10.1109/CICT.(**2016**), pp.142.

[5] M. Services, Medicare Fee-For Service Provider Utilization and Payment Data Inpatient Public Use File : A Methodological Overview **(2016)**.

[6] T. Santhanam, Comparison of K-Means Clustering and Statistical Outliers in Reducing Medical Datasets, **(2014)**, pp. 1–6.

[7] R. Paul, A. Sayed and L. Hoque, Clustering Medical Data to Predict the Likelihood of Diseases, **(2010)**, pp. 44–49.

[8] M. Okuda, The change of the structure of patient satisfaction by waiting time, length of stay and hospital rebuilding in Japan, **(2015)**, 336–344, accessed from https://doi.org/10.1109/ICDMW.2015.238.

[9] S. C. Morajkari and J. A. Laxminarayani, Threshold Based Similarity Clustering of Medical Data, **(2014)**, (978), pp. 591–595.

[10] A.M. Mehar, A. Maeder, K. Matawie and A. Ginige, Blended Clustering for Health Data Mining, **(2010)**, pp. 130–137.

[11] E. Mathematics, (n.d.), Possibilistic fuzzy c-means clustering on medical, (1), pp. 2–6.

[12] A. Mahmood, K. Shi and S. Khatoon, Controlling In-patient Environment by Mining Sensor Data, **(2013)**, pp. 674–679.

[13] S. Li, X. Zhou, H. Shi and Z. Zheng, An Efficient Clustering Method for Medical Data Applications, **(2012)**, pp. 133–138.

[14] K. Lemke, A Predictive Model to Identify Patients at Risk of Unplanned 30-Day Acute Care Hospital Readmission, **(2013)**, accessed online from 551–556. https://doi.org/10.1109/ICHI.2013.86.

[15] C. Klüver, Steering Clustering of Medical Data in a Self- Enforcing Network (SEN) with a Cue Validity Factor, **(2016)**.

[16] S. Khanna, J. Boyle, N. Good, J. Lind and K. Zeitz, Time Based Clustering for Analyzing Acute Hospital Patient Flow, **(2012)**, pp. 5903–5906.

[17] S. Khanmohammadi, N. Adibeig and S. Shanehbandy, An improved overlapping k-means clustering method for medical applications. Expert Systems with Applications, *67*, 12–18 **(2017)**, accessed online from https://doi.org/10.1016/j.eswa.2016.09.025.

[18] TanviAnand, Rheka Pal and S.K. Dubey, Data Mining in Healthcare Informatics. Techniques and Applications, **(2016)**, pp. 4023–4029.

[19] A.K. Hmood and B.C.M. Fung, Privacy-Preserving Medical Reports Publishing for Cluster Analysis **(2014)**.

[20] P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas and N. Matsatsinis, Knowledge-Based Systems supporting healthcare management decisions via robust clustering of event logs. Knowledge-Based Systems, **(2015)**, accessed online from https://doi.org/10.1016/j.knosys.2015.04.012, pp. 203–213

[21] J.H. Chen, M. Alagappan, M.K. Goldstein, S. M Asch and R. B. Altman, International Journal of Medical Informatics Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets. International Journal of Medical Informatics, **(2017)**, pp. *102*, 71–79, accessed online from https://doi.org/10.1016/j.ijmedinf.2017.03.006.

[22] T. Cerquitelli, S. Chiusano and X. Xiao, Exploiting clustering algorithms in a multiple-level fashion : A comparative study in the medical care scenario, **(2016)**, accessed online from https://doi.org/10.1016/j.eswa.2016.02.005.

[23] M. Bhattacharya, D.Ehrenthal and H. Shatkay, ,Identifying Growth-Patterns in Children by Applying Cluster analysis to Electronic Medical Records, **(2014)**, pp. 348–352.

Authors

Kerina Blessmore Chimwayi, she did her B.Tech in Electronic Commerce from Harare Institute of Technology, Harare, Zimbabwe and is currently pursuing M.Tech in Computer Science & Engineering (with Specialization in Big Data Analytics) at VIT University, Vellore. Her research interests include Machine Learning, Deep Learning and Spatial Data Mining in Healthcare.

Noorie Haris, she did her B.tech in Computer Science and Engineering from MES College of Engineering, Kuttipuram, Kerala and is currently pursuing M.tech in Computer Science and Engineering with specialization in Big Data Analytics at VIT University, Vellore. Her research interests include advanced analytics using big data technologies.

Ronnie D. Caytiles, he had his Bachelor of Science in Computer Engineering- Western Institute of Technology, Iloilo City, Philippines, and Master of Science in Computer Science– Central Philippine University, Iloilo City, Philippines. He finished his Ph.D. in Multimedia Engineering, Hannam University, Daejeon, Korea. Currently, he serves as an Assistant Professor at Multimedia Engineering department, Hannam University, Daejeon, Korea. His research interests include Mobile Computing, Multimedia Communication, Information Technology Security, Ubiquitous Computing, Control and Automation

N. Ch. S. N. Iyengar (b 1961), currently Professor, Information Technology, Sreenidhi Institute of Science (SNIST) and Technology, Yamnapet, Ghatkesr, Hyderabad-501301, Telengana, India He is a doctorate in both Applied Mathematics and Computer Science and Engineering. His research interests include Agent-Based Distributed Computing, Intelligent Computing, Network Security, Cloud Computing, Big Data Analytics and Fluid Mechanics. He had 32+ years of experience in teaching to B.Tech. and M.Tech students. He guided 12 Ph.Ds, 5 M.Phils and 75 + M.Tech Projects apart from authoring several textbooks. He had 200+ research publications in reputed peer reviewed international journals along with students. He organized many conferences/workshops and continuing education programmes He served as Keynote speaker/ / Invited speaker //PCM/reviewer for many International conferences. He serves as a Editor in chief/Guest editor /Editorial board member for many international journals. He is the professional member of many bodies.