# Data Mining Classification Methods for Pediatric Records of Fujairah Hospital

Abdelaziz Araar, Rashid Alamiri

College of Information Technology,
Ajman University, UAE,

## *Abstract*

*More than half of the children in the UAE will be prone to fatal chronic diseases in the next few years due to obesity and other behaviors, and the situation needs to be tackled urgently. Data mining on medical data has great potential to improve the treatment quality of hospitals and increase the survival rate of patients. Early prediction techniques have become an apparent need in many clinical areas.*

*We propose two-phase approach; the first phase uses data mining techniques with classification methods. The dataset collected has predefined classes. Among 100 children in Fujairah hospital, we modeled with 43 attributes and 16 classes. We used Decision Trees (J48), Rules Induction (PART), Bayesian Network (BayseNet), Multilayer Perceptron (MLP), and Nearest Neighbor (NNge) to find rules and conclusion. In the second phase, we establish a comparison among the five classifiers. Empirical results showed that the developed models could classify diseases within reasonable accuracy. The comparison of these classifiers showed that the NNge classifier is the best classifier for most of all classes. In this paper, we used a clustering technique to identify patient outliers who are using a large amount of drugs over a period of time. Finally, we generated recommendations and conclusions.*

***Keywords:*** *Data mining, PART, BayesNet, MLP, NNge, Pediatric , and Drug Utilization.*

## I. INTRODUCTION

This Data mining on medical data has great potential to improve the treatment quality of hospitals and increase the survival rate of patients. Clinical study has found early detection and intervention to be essential for preventing clinical deterioration in patients at general hospital units. High rates of obesity among school children are putting them at risk of developing heart disease and diabetes at a younger age, a study has found. Dubai Thalassemia Center organized a campaign to help patients suffering from excessive iron levels.

Researchers at UAE University studied 1,018 students age 12 to 18, measuring their cholesterol levels and blood pressure as well as height and weight. Just over half of the sample was Emirati. The study also found that nearly 30 per cent of obese children had elevated blood pressure, compared with 20 per cent of overweight children and 8 per cent of those with normal weight [1].

Data mining on electronic medical records has attracted a lot of attention but is still at an early stage in practice. Knowledge of such subgroups of patients is valuable for tailoring and implementing quality initiatives in hospitals. The growing adoption of information technologies in healthcare and the availability of more patient data and related healthcare variables provide new opportunities for using analytics to impact health outcomes. Most hospitals have some type of healthcare information technology in place (e.g. Electronic Medical Record, EMR), which allows them to measure and capture patient data in real time and at the point the care [2].

### A. Importance of Healthcare

These are the some important features in Healthcare [3].

- Access all the patient records and rapidly detect anomalies,
- Analyze data using an automated system, which is useful in the case of major and repeated anomalies.
- Boost productivity and care quality through remote, shorter and more frequent consultations.
- Interact quickly and easily in a structured way via tools shared between the primary care provider and day-to-day patient monitoring,
- Provide motivational support for patients who desire it,
- Contribute to biomedical research through the tool‘s healthcare database.

The objective of this study is to provide useful rules and compare different classifiers.

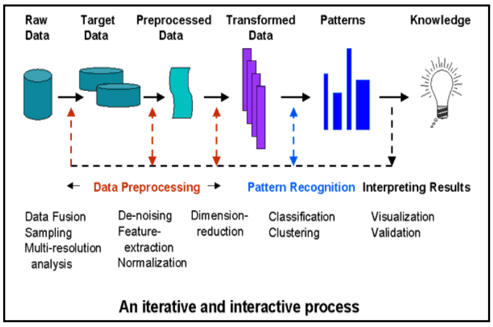


Figure 1: An Overview of the Steps that Compose the KDD Process.

The above figure is used to analyze and interpret these data. Moreover, the paper proposes a two-phase approach to solve the above difficulty.

In phase one, we use Decision Trees (J48), Rules Induction (PART), Bayesian Network (BayseNet), MultilayerPerceptron (MLP), and Nearest Neighbor (NNge) to find rules and conclusion [4]. In phase 2, we establish a comparison among the classifiers.

## II. RELATED WORK

Process mining was originally conceived to analyze and visualize event logs from non-healthcare related industries, and so it is not built for handling specific relationships (e.g., medication—outcome) that are seen in healthcare, but process-mining related techniques for visualization may inform our own future techniques for clinical data analysis [5].

Many studies combine process mining with other data mining techniques, such as clustering methodologies based on Hidden Markov Model transition matrices and k-means clustering to identify regular, infrequent, and outlier clinical cases [6], as well as spectral clustering to identify the major activity patterns in an emergency department [7]. Other studies that apply process mining to healthcare are reviewed elsewhere [8].

Few studies related to medical diagnosis and survivability using data mining approaches like decision trees [10]. In [11], they preprocessed the SEER data for to remove redundancies and missing information. The resulting data set had 202,932 records, which then pre-classified into two groups of ―survived” (93,273) and ―not survived” (109,659) depending on the Survival Time Recode (STR) field. The ―survived” class is all records that have a value greater than or equal 60 months in the STR field and the ―not survived” class represent the remaining records. After this step, the data mining algorithms are applied on these data sets to predict the dependent field from 16 predictor fields. The results of predicting the survivability were in the range of 93% accuracy. After a careful analysis of the breast cancer data used in [11], we have noticed that the number of ―not survived” patients used does not match the number of ―not alive” (field VSR) patients in the first 60 months of survival time. As a matter of fact, the number of ―not survived” patients is expected to be around 20% based on the breast cancer survival statistics of 80% [12]. In [13], they used decision trees to extract clinical reasoning in the form of medical expert‘s actions that are inherent in a large number of electronic medical records. The extracted data could be used to teach students of oral medicine a number of orderly processes for dealing with patients with different problems depending on time. In [14], they utilized a C4.5 algorithm to build a decision tree in order to discover the critical causes of type II diabetes. She has learned about the illness regularity from diabetes data, and has generated a set of rules for diabetes diagnosis and prediction. In [15], they discovered _treatment pathways‘ through mining medical treatment procedures in the emergency department. They found that the workload in the emergency department varies depending on the number of presented patients, and is not affected by the type of procedure carried out. In [16], she has presented a complementary perspective on the activities of the emergency department for specific patient groups: over 75 year old and under 75 years old patients. She thought once validated, these views would be used as decision support tools for delivering better care to this population. In [17], they found a way to raise the accuracy of triage through mining abnormal diagnostic practices in the triage. A two-stage cluster analysis (Ward‘s method, K-means) and a decision tree analysis were performed on 501 abnormal diagnoses done in an emergency department.

In this paper, we use real pediatric data from the Fujairah hospital. WEKA is used as a tool to generate results with different classification techniques. Suggestions and recommendations are derived.

## III. EXPERIMENTAL SETUP

### *A. Data Collection and Preprocessing*

Figure below shows the admission and discharge processes of the hospital for the pediatric department.

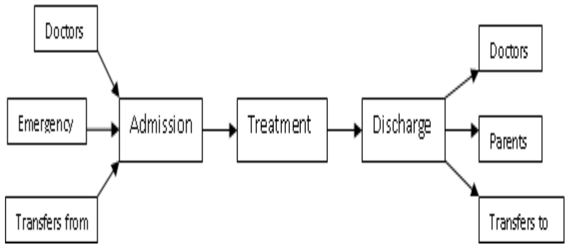


Figure 2: Hospital Process

Data Preprocessing is an essential part of Data Preparation. Different methods of data preprocessing were used to the patient‘s data in this paper. Data Preprocessing involved many tasks such as data aggregation, feature selection and creation, and normalization.

The cleaned dataset contains 43 attributes and has a predefined class. Attributes are created based on the real data which has been collected as shown in Table 1. Dataset is formatted to the ARFF (Attribute-Relation File Format) to describe a list of instances sharing a set of attributes. Tables 2, 3, 4 and 5 describe the nationality, the group of medicines, the symptoms, and the diagnostics respectively.

TABLE 1 ATTRIBUTES DECLARATION AND DESCRIPTION

| No | Attribute name | Type | Description |
|---|---|---|---|
| 1 | Patient_SL | Numeric | |
| 2 | Age | Numeric | |
| 3 | Nationality | Numeric | See Table 2 |
| 4 | Gender | {0,1} | |
| 5 | Duration | Numeric | From entry to exit |
| 6 | admission_emergency | Numeric | |
| 7 | admission_doctor | Numeric | |
| 8 | admission_transfer | Numeric | |
| 9 | discharge_med_advice | Numeric | |
| 10 | discharge_parent | Numeric | |

| | | | |
|---|---|---|---|
| 11 | discharge_transfer | Numeric | |
| 12-37 | Symptoms | {0,1 } | See Table 4 |
| 38 | Med_Pain_killers | {0,1 } | See Table 3 |
| 39 | Med_Antibiotics | {0,1 } | |
| 40 | Med_Vitamins | {0,1 } | |
| 41 | Med_Chronic_ diseases | {0,1 } | |
| 42 | Med_Treatments | {0,1 } | |
| 43 | 16 Classes | Nominal | See Table 5 |

The table below contains 12 codes of nationality for the patients.

TABLE 2: NATIONALITY AND CORRESPONDING CODE

| Code | Nationality | Code | Nationality |
|---|---|---|---|
| 1 | UAE | 7 | No papers |
| 2 | GCC | 8 | African |
| 3 | Yemen | 9 | Middle East |
| 4 | India | 10 | North Africa |
| 5 | Pakistan | 11 | Europe |
| 6 | Iran | 12 | Others |

The table below contains 5 categories of medicines

TABLE 3: ATTRIBUTES DECLARATION AND DESCRIPTION

| No | Medicine-type | Medicine |
|---|---|---|
| 1 | Med_Pain_killer | Paracetamol, Buscopane, Ibuprofen |
| 2 | Med_Antibiotics | Augmentin, Amoxicillin, Ampicillin, Amikacin, Meropeum, Gentamicin, Azithromycin, Cefuroxime, Clarithromycin, Cefotaxime, Cefixime, Gentamicin, Cefprozil |
| 3 | Med_Vitamins | Vitamin, ORS, Dextrose IVF, |
| 4 | Med_Chronic_ disease | Insulin, Baclofen, Ventolinneb, Diazepam, Montelukas, Levetiracetam, IV immunoglobuline, Ondansetron |
| 5 | Med_Treatment | Blood Transfusion, Atrovent, Chlorpheniramine, Prednisolone, Ranitidine, Ceftriaxone, Loratidine, Phototherapy, IV Rocephine, Scopolamine, Adrenaline, Diphenhydramine, Domperidone, Monteluka |

The table below contains 26 symptoms for different diseases.

TABLE 4: DIFFERENT SYMPTOMS

| No | Symptoms | No | Symptoms |
|---|---|---|---|
| 12 | High_Fever | 25 | Weakness |
| 13 | Vomiting | 26 | Weight_loss |
| 14 | Chest pain | 27 | Frequent_urination |
| 15 | Abdominal_pain | 28 | Excessive_thirst |
| 16 | Caught | 29 | Dizziness |
| 17 | Diarrhea | 30 | Yellowing_eyes |
| 18 | Chills | 31 | Loss_of_ appetite |
| 19 | Ear_pain | 32 | Seizures |
| 20 | Rash | 33 | Lack_of_urine |
| 21 | Runny nose | 34 | Low_blood_sugar |
| 22 | Wheezing | 35 | Slow _or_fast_ heart_rate |
| 23 | Difficult_or_rapid_ breathing | 36 | Throat_pain |
| 24 | Headache | 37 | Anuria |

The table below contains 16 causes of admission

TABLE 5: CLASSES/CAUSES OF ADMISSION

| Code | Diagnostics (classes) |
|---|---|
| 1 | Pneumonia |
| 2 | Gastroenteritis |
| 3 | Bronchiolitis |
| 4 | Jaundice |
| 5 | *Upper respiratory tract infections (URTI)* |
| 6 | Asthma |
| 7 | Dehydration |
| 8 | Hepatitis |
| 9 | Epilepsy (Febrile seizure) |
| 10 | Diabetes (IDDM) |
| 11 | *Henoch-Schönlein purpura (HSP)* |
| 12 | *Cystic fibrosis (CF)* |
| 13 | *Neonatal sepsis* |
| 14 | Pharyngitis |
| 15 | *Septicemia* |
| 16 | Tonsillitis |

### *B. WEKA Toolkit*

The *WEKA* (Waikato Environment for Knowledge Analysis) is an easy to use graphical user interface that harnesses the power of the WEKA software [18]. The major WEKA packages are Filters, Classifiers, Clusters, Associations, and Attribute Selection is represented in the Explorer along with a Visualization tool, which allows datasets and the predictions of Classifiers and Clusters to be visualized in two dimensions.

### *C. Simplified methodology*

Classification is a process of finding a model that describes and distinguishes data classes or concepts, for the purpose arranging the records into different class labels and also to predict the class of objects whose class label is unknown.

Using WEKA tool‘s preprocessing and classification methods we classified the disease records in pediatric department into different types. This classification lets us to predict some needed results for our study. Five classification methods; decision trees J48 [12],[20], rules PART [21], naive bayes [21], neural nets [22],[23] and NNge [24] are applied on the dataset.

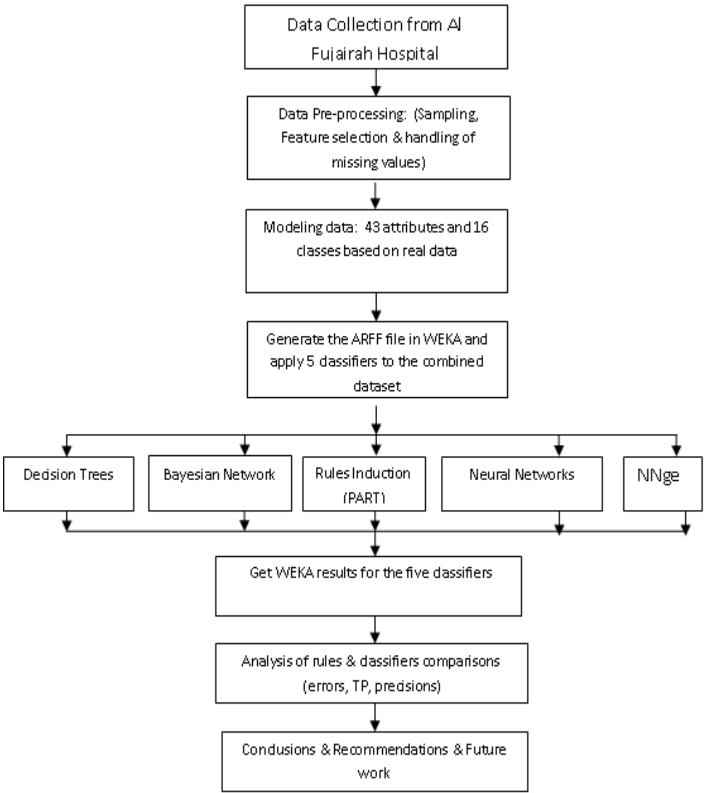
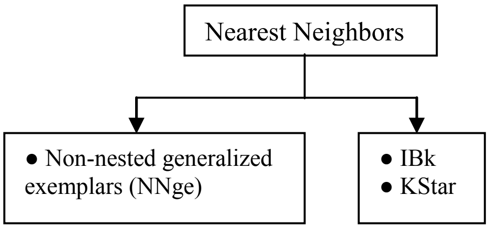

Data Collection from Al Fujairah Hospital
Data Pre-processing: (Sampling, Feature selection & handling of missing values)
Modeling data: 43 attributes and 16 classes based on real data
Generate the ARFF file in WEKA and apply 5 classifiers to the combined dataset
Decision Trees
Bayesian Network
Rules Induction (PART)
Neural Networks
NNge
Get WEKA results for the five classifiers
Analysis of rules & classifiers comparisons (errors, TP, precisions)
Conclusions & Recommendations & Future work

Figure 3: Simplified Methodology of the process

## IV. EXPERIMENT RESULTS

### *A. Phase 1: Results*



Nearest neighbor classifiers are popular kind of lazy learning algorithms which are based on learning by analogy, that is, by comparing a given test instance with training instances that are similar to it. NNge algorithm using non-nested generalized exemplars is hyper-rectangles that can be viewed as if-then rules [24].

The method provides 16 rules which cover all attributes in the dataset. Some classes have only one rule per class, but other may have many rules per class. We show only one rule for space constraint.

*RULE 1*

**IF** : 3.0<=Patient_SL<=100.0 ^ 35.0<=Age<=3285.0 ^
0.0<=Gender<=1.0 ^ 1.0<=Nationality<=12.0 ^
0.0<=duration<=4.0 ^ 0.0<=discharge_med_advice<=1.0 ^
0.0<=discharge_parent<=1.0 ^ discharge_transfer=0.0 ^
0.0<=admission_emergency<=1.0 ^
0.0<=admission_doctor<=1.0 ^ 0.0<=admission_transfer<=1.0 ^
High_Fever=1.0 ^ Vomiting=1.0 ^ Chest_pain=0.0 ^
Abdominal_pain=1.0 ^ Caught=0.0 ^ Diarrhea=1.0 ^ Chills=0.0
^ Drowsy=0.0 ^ Rush=0.0 ^ Runny_nose=0.0 ^
Wheezing=0.0 ^ Difficult_breathing=0.0 ^ Headache=1.0
^ Weakness=0.0 ^ Weight_loss=0.0 ^
frequent_urination=0.0 ^ Excessive_thirst=0.0 ^
Tiredness=0.0 ^ Yellowing_eyes=0.0 ^ Loss_appetite=0.0
^ Seizure=0.0 ^ Lack_urine=0.0 ^ Low_blood_sugar=0.0
^ Slow_or_fast_heart_rate=0.0 ^ Rapid_breathing=1.0 ^
Anuria=0.0 ^ 0.0<=Med_Pain_killer<=1.0 ^
0.0<=Med_Antibiotics<=1.0 ^ 0.0<=Med_Vitamins<=1.0
^0.0<=Med_Chronic_disease<=1.0 ^
0.0<=Med_Treatment<=1.0 **THEN** Class Gastroenteritis

Interpretation of rule 1:
Based on 29 cases, the patient could have gastroenteritis disease if:
the maximum duration is 4 days AND he/she has high fever, vomiting, abdominal pain, diarrhea, headache, rapid breathing AND taking pain killers with or without other medicines.

### *B. Phase 2: Comparison of classifiers*

The following table shows the accuracy of classifiers:

TABLE 6: ACCURACY OF CLASSIFIERS

| **Classifier`s Name** | **Correctly classified instances** | **Accuracy in percentage** |
|---|---|---|
| Trees (48) | 92 | 92 % |
| Bayes (BayesNET) | 94 | 94 % |
| Rule (PART) | 87 | 87 % |
| MLP | 90 | 90 % |
| NNge | 90 | 90 % |

*Let TP, FP, FN and TN*, denote true positive rate, false positive rate, false negative and true negative respectively [20].

TABLE 7: SUMMARY OF MEASURES

| | | Predicted class(observation) | |
|---|---|---|---|
| | | symptom | No symptom |
| Actual expectation | disease | TP : correct result | FN (Type II error) |
| | No disease | FP: (type I error) | TN: correct absence of results |

Positive predicted value (Precision) = TP/(TP+FP)
Accuracy = (TP +TN) /( TP + TN+ FP + FP)

Accuracy indicates proximity of measurement results to the true value; it is from the confusion matrix.
The measurement system is designated valid, if it is both accurate and precise. Classifiers with high precision and accuracy and low FP rates are preferred. In this paper, we use a new measurement:

$$AP = accuracy * precision \quad (1)$$

1= Trees, 2 = Bayes, 3= Rules, 4= MLP, and 5 = NNge.

TABLE 8: THE PRODUCT OF TP RATE BY PRECISION FOR ALL CLASSIFIERS

| **Classes** | TP rate * Precision | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Pneumonia | 0.86 | 0.9 | 0.86 | 0.86 | 0.95 |
| Gastroenteritis | 0.9 | 1 | 0.93 | 0.96 | 1 |
| Bronchiolitis | 1 | 1 | 1 | 0.92 | 1 |
| Jaundice | 1 | 0.75 | 0.44 | 0.67 | 0.44 |
| *(URTI)* | 0.9 | 0.9 | 0.9 | 0.71 | 0.95 |
| Asthma | 0.56 | 0.57 | 0.33 | 0.45 | 0 |
| Dehydration | 0.7 | 0.83 | 0.44 | 0.84 | 0.46 |
| Hepatitis | 0 | 0 | 0 | 0 | 0 |
| Epilepsy | 1 | 0.95 | 0.6 | 0.75 | 1 |
| Diabetes | 1 | 0.67 | 0.5 | 1 | 0.7 |
| *HSP* | 1 | 1 | 0.67 | 1 | 1 |
| *CF* | 0 | 0 | 0 | 0 | 0 |
| *Neo* | 0 | 0 | 0 | 0 | 0 |
| Pharyngitis | *0* | *1* | *0* | *0* | *1* |
| *Septicemia* | 0 | 0 | 0 | 0 | 0 |
| Tonsillitis | *1* | *1* | *1* | *1* | *1* |

Table 8 and figure 4 show the comparison among these classifiers based on AP. The impact of AP is amplified in order to conclude which algorithm is the most suitable one for our application. For our study, NNge classifer has a higher performance measure.

| Classifiers | Tree | BN | PART | MLP | NNge |
|---|---|---|---|---|---|
| AP frequency / 16 | 6 | 6 | 2 | 4 | **8** |

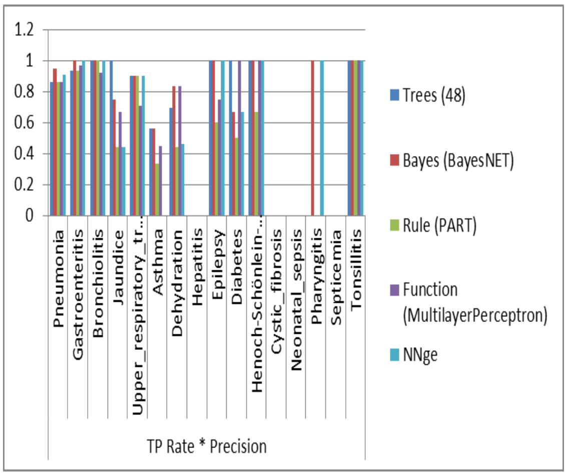


Figure 4: The product of TP Rate by Precision for all classifiers

False positive rate FP = TN / (TN +FN)
FP Rate is shown figure 5 to give a better picture for each class.

| Classifiers | Tree | BN | PART | MLP | NNge |
|---|---|---|---|---|---|
| FP frequency / 16 | 12 | 12 | 10 | 10 | 12 |

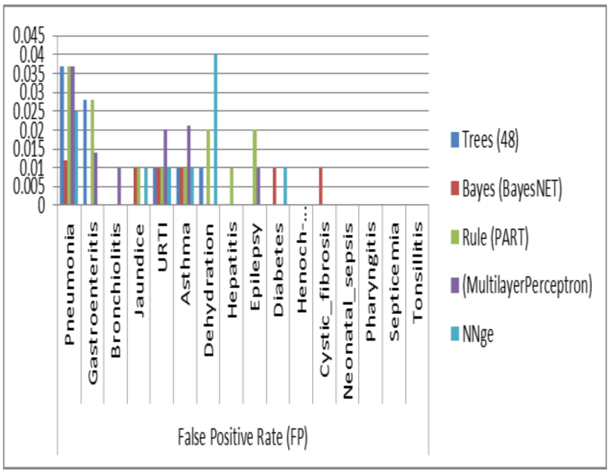


Figure 5: the FP Rate for all classifiers

Regarding the FP rate frequency, 3 classifiers have better values for 12 classes among 16, including NNge classifier. Now, we use K-mean as a clustering method to find patient outliers [25] . We remove classes from the dataset, and then we use WEKA. This process is used to compare the same outliers against hospital diagnostic outliers.

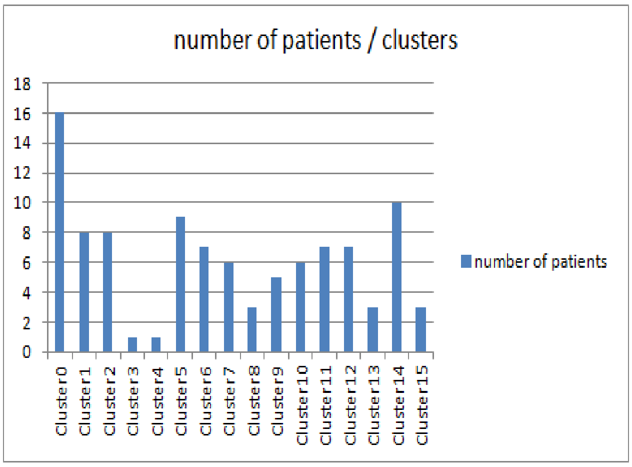


Figure 6: K-mean clustering with K = 16.

From figure 6, we found that the outliers are in clusters 3 and 4. Patient numbers 15 and 46 are outliers in the dataset according to SimpleKmeans. Regarding the diagnostic, patient numbers 15, 39, 46, and 57 have a unique disease. However, K-mean captures only 50 % of the outliers.

## V. CONCLUSION & FUTURE WORKS

In this paper, the focus is on the analysis of pediatric records in Fujairah hospital and the investigation of the performance of five classification algorithms. From empirical results, it was found that NNge could offer sufficient insight into the dataset being studied.
Based on the review process of child disease reports and obtained results, some recommendations related to pediatric are suggested to improve control and safety in

Fujairah hospital. The following are some of these recommendations:

- Early disease detection and regular checking for detecting like; diabetes, thalassemia, etc.
- Reducing exposure to environmental tobacco smoke
- Reducing Exposure to Cockroaches
- Reducing Exposure to Pets
- Improve diet and living area
- The Growth of Obesity and Technological Change
- More awareness campaigns should be conducted related to the distraction attentions.

Future work includes improvement of the accuracy of the current result as well as further utilization of recognized news and comment messages**.** As future work, this work can also be extended to:

1. More number of records for a period of 3 to 5 years for better analysis.
2. Data mining for other hospitals in UAE,
3. Text mining system can help decision makers when considering whether or not scans may be helpful in reaching a diagnosis.

## ACKNOWLEDGMENT

We would like to thank the Director of the Fujairah Hospital, Mr. Obaid Khalfan Al Khadim for providing the data.

## REFERENCES

[1] Obese UAE children at risk of early heart disease, The National, UAE newspaper , May 2012.

[2] S. Espinoza, ―Data Mining Methods to healthcare problems", Ph.D. Dissertation, G.I.T., USA, 2012

[3] M. Dey, S. S. Rautaray, ―Study and Analysis of Data Mining Algorithms for Healthcare Decision Support System", International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 470-477, 2014

[4] A. Araar and A.A. El Tayeb, ―Mining Road Traffic Accident Data to Improve Safety in Dubai", Journal of Theoretical and Applied Information Technology, Vol 47, No 3, pp 911- 925, 2013

[5] W. M. P. van der Aalst, ―Process Mining: Discovery, Conformance and Enhancement of Business Processes", Springer Publishing Company, Incorporated, 1st edition, 2011.

[6] A. Rebuge and D. R. Ferreira, ―Business process analysis in healthcare environments: A methodology based on process mining", Inf. Syst., 37(2):99–116, Apr. 2012

[7] P. Delias, M. Doumpos, P. Manolitzas, E. Grigoroudis, and N. Matsatsinis, ―Clustering Healthcare processes with a robust approach", In 26th European Conference on Operational Research, 2013.

[8] R. Mans, W. V. Aalst, R. Vanwersch, and A. Moleman.―Process mining in healthcare: Data challenges when answering frequently posed questions", Springer Berlin Heidelberg, volume 7738 of Lecture Notes in Computer Science, pages 140–153. 2013

[9] R. Bellazzi and B. Zupan. Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 77(2):81– 97, 2008.

[10] R. Ceglowski, L. Churilov, and J. Wasserthiel, ―Combining data mining and discrete event simulation for a value-added view of a hospital emergency department," Journal of the O. R. Society, vol.58, pp. 246-254, 2007

[11] C. Duguary, and F. Chetouane, ―Modeling and Improving Emergency Department Systems using Discrete Event Simulation", Simulation, vol. 83, pp. 311-320, 2007

[12] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive information visualization to explore and query electronic health records. Foundations and Trends in Human-Computer Interaction, 5(3):207–298, 2013.

[13] K. H. Butler and S. A. Swencki, ―Chest pain: a clinical assessment," Radiologic Clinics of North America, vol. 44, pp. 165-179, 2006.

[14] Y. Tan, G. F. Yin, G. B. Li, and J. Y. Chen, ―Mining compatibility rules from irregular Chinese traditional medicine database by Apriori algorithm," Journal of Southwest JiaoTong University, vol. 15, 2007.

[15] H. Ren, ―Clinical diagnosis of chest pain," Chinese Journal for Clinicians, vol. 36, 2008

[16] B. Riccardo and Z. Blaz, ―Predictive data mining in clinical medicine: Current issues and guidelines," International Journal of Medical Informatics, vol. 77, pp. 81- 97, 2008

[17] Y. P. Yun, ―Application and research of data mining based on C4.5 Algorithm," Master thesis, Haerbin University of Science and Technology, 2008

[18] www.cs.waikato.ac.nz/aml/weka

[19] Ashokkumar Vijaysinh Solanki , ―Data Mining Techniques Using WEKA classification for Sickle Cell Disease", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 5857- 5860, 2014

[20] I. Moghimipour, M. Ebrahimpour, ―Comparing Decision Tree Method Over Three Data Mining Software", International Journal of Statistics and Probability; Vol. 3, No. 3; 2014.

[21] Xhemali D, Hinde C J & Stone R, ―Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web pages", IJCSI V4 N1, pp 16-23, 2009

[22] Trilok Chand Sharma and Manoj Jain, ―WEKA Approach for Comparative Study of Classification Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, pp 1925- 1932, 2013.

[23] P.Ramachandran, .N.Girija, T.Bhuvaneswari, ―Cancer Spread Pattern – an Analysis using Classification and Prediction Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 6, June 2013.

[24] P. Kulkarni, R. Ade, ―Prediction of Student's Performance based on Incremental Learning", International Journal of Computer Applications, V. 99 – No.14, 2014

[25] M.S. Anbarasi, S. Ghaayathri, R. Kamaleswari,and I. Abirami, Outlier Detection for Multidimensional Medical Data, International Journal of Computer Science and Information Technologies, Vol. 2 (1) , 512-516, 2011.