Intelligent Data Analysis: the Best Approach for Chronic Heart Failure (CHF) Follow Up Management

Niloofar Mohammadzadeh1, Reza Safdari2, Alireza Baraani3, Farshid Mohammadzadeh4

Health Information Management, Tehran University of Medical Sciences. Tehran, Iran
Health Information Management Department, Health Information Management Research Center, Tehran University of Medical Sciences, Tehran, Iran
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
Shahabe Danesh University, Iran

Corresponding author: Reza Safdari, PhD, Associate Professor, Head of Health Information Management Department, Deputy of health information management research center, Tehran University of Medical Sciences, Tehran, Iran. Tel: +98-21-88951922, Fax: +98-21-88607963, E-mail: nmohammadzadeh@razi.tums.ac.ir

ABSTRACT
Objective: Intelligent data analysis has ability to prepare and present complex relations between symptoms and diseases, medical and treatment consequences and definitely has significant role in improving follow-up management of chronic heart failure (CHF) patients, increasing speed and accuracy in diagnosis and treatments; reducing costs, designing and implementation of clinical guidelines. The aim: The aim of this article is to describe intelligent data analysis methods in order to improve patient monitoring in follow and treatment of chronic heart failure patients as the best approach for CHF follow up management. Methods: Minimum data set (MDS) requirements for monitoring and follow up of CHF patient designed in checklist with six main parts. All CHF patients that discharged in 2013 from Tehran heart center have been selected. The MDS for monitoring CHF patient status were collected during 5 months in three different times of follow up. Gathered data was imported in RAPIDMINER 5 software. Results: Modeling was based on decision trees methods such as C4.5, CHAID, ID3 and k-Nearest Neighbors algorithm (K-NN) with k=1. Final analysis was based on voting method. Decision trees and K-NN evaluate according to Cross-Validation. Conclusion: Creating and using standard terminologies and databases consistent with these terminologies help to meet the challenges related to data collection from various places and data application in intelligent data analysis. It should be noted that intelligent analysis of health data and intelligent system can never replace cardiologists. It can only act as a helpful tool for the cardiologist’s decisions making.

Key words: Data Analysis, Heart Failure, Follow Up Study.

1. INTRODUCTION
Chronic heart failure (CHF) is a major public health problem in most countries (1). In Europe, CHF is the leading cause of elderly hospitalization and 1-2% of all hospitalizations due to this disease (2). Total heart failure costs are estimated at 44.6 billion in 2015 only in the United States (3). Symptoms and morbidity associated with CHF effect patients’ daily activities and quality of life. High rate of heart failure readmission in early months after discharge are common. So finding the opportunities to reduce readmission and improve the quality of patients life is very essential (4). CHF disease management programs especially follow up plans with patient center approach can improve survival, quality of life and decrease re hospitalization. Also note treatment burden of patients in follow up plan is very important. Patients’ needs to adapt to new life style changes, learn about self-management strategies, engage with others, monitor the treatments, and integrate treatments into social circumstances and overcoming barriers such as accessibility to health care and poor continuity of care (5, 6).

Information technology tools like telemedicine and tele care systems use effectively to provide require services for chronic disease patients. These systems would be helpful to improve the monitoring of patients through continuous assessment of symptoms and signs of disease and checking compliance with self-management programs, thus they can improve treatment, and resource usage (7, 8).

Use of intelligent systems lead to facilitate, accelerate and improve health care services, decision support in the areas of diagnosis and treatment particularly in home care and telemedicine. Data analysis with artificial intelligence methods are important tools that can provide access to hidden knowledge in large databases of clinical and administrative data in hospitals. Also improve decision support, prevention, treatment and diagnosis planning and provides useful knowledge for health care personnel (9).

Intelligent data analysis has ability to prepare and
present complex relations between symptoms, diseases, and medical and treatment consequences (10). It is obvious that intelligent data analysis definitely have significant role in improving health care, prevention, treatment, and follow-up management of CHF patients, increase quality of health care, high speed and accuracy in diagnosis and treatment, reduce costs, clinical outcomes assessment, design and implementation of clinical guidelines (11). The aim of this article is to describe intelligent data analysis method in order to improve patient monitoring in follow and treatment of chronic heart failure patients as the best approach for CHF follow up management.

2. METHODS

In this research according to standard guideline about diagnosis and management of chronic heart failure in America (12), Europe (13), Australia (14) and consult with a cardiologist, minimum data set requirements for monitoring and follow up of CHF patient were designed in checklist with six main parts include demographic, clinical, examination, exercise test, drug, procedure and imaging. These minimum data sets were distributed among 16 cardiologists and their comments for each item were collected after discussion group. All CHF patients that discharged in 2013 from Tehran heart center have been selected. Then MDS for monitoring CHF patient status were collected during 5 months in three different times of follow up. Gathered data was imported in RAPIDMINER 5 software for intelligent data analysis.

3. RESULTS

Important factors that studied in this research for follow up CHF patients include:

- Demographic: age, sex, hypertension, LDL, diabetes, cigarette, family history;
- Clinical: chest pain, PND, dyspnea, edema;
- Examination: NA, K, HB, BUN, CR;
- Exercise Test: functional capacity, ST Low, chest pain during exercise test;
- Drugs: diuretics, digoxin, statins, aspirin, CA blockers, ace inhibitors, ARB, beta blockers, nitrates 6-Imaging / Procedure include: a) Angiography: coronary artery involvement; b) Echo:EF, left ventricular size, mitral valve regurgitation, pulmonary artery pressure, right ventricular function; c) Scan SPECT:ischemia, viability and (d) ECG:ECG normality, ST-T abnormality, LVH, LA abnormal.

From total amount of patients in the sample 44 percent of chronic heart failure patient (36 people) were female and 56 percent (48 people) were male.12percent (10 people) were 17-49 year old, 46percent (37 people) were 50-65 years and 42 percent (34 people) were aged over 66 years. 54 percent (44 people) had blood pressure, 40 percent (32 people) had diabetes, 53 percent (43 people) had decrease hemoglobin, and 75 percent (61 people) had hyper urea (BUN).

In order to manage missing value in this project, records with most blank values or missing values were excluded from the intelligent data analysis. Details of some of the follow up data were not recorded in information system in Tehran heart center, so these variables had not positive effect on the data analysis. Because of this, data analysis performed based on variables which information was to be recorded in the hospital information system and other attributes were deleted. Data discretization was performed after data cleansing and preprocessing process in order to provide suitable basis for more efficient performance of learning model algorithms. Attributes that selected for discretization were show in Table1.

Modeling was based on decision trees methods such as C4.5, CHAID, Id3 with k=10 and k-Nearest Neighbors algorithm (K-NN) with k=1. Final analysis was based on voting method. Decision trees and K-NN evaluate according to Cross-Validation. Cross-Validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. This method data dived into two parts (train and test) and data in both parts are repeated k times and assessed regularly (15).

All relevant data of first follow up was attached to set role function in RAPIDMINER 5 and patient status in second follow up includes follow up or non-follow up (death, hospitalization) with this function as the label (feature target) was considered. Then necessary changes with select attribute and rename

| Attribute         | Indicator                  | Attribute          | Indicator                  |
|-------------------|----------------------------|--------------------|----------------------------|
| Age               | Young=17-49, adult= 50-65, old=66year | EF                 | EF=30, severe, EF =45, mild |
| LDL               | normal;;130, abnormal=130  | K                  | K=3.5                      |
| K                 | “decrease”, K=5 "normal", K=3 "hyper" | Pulmonary Artery Pressure | <<25 "normal", >25 "hyper" |
| Hb                | Hb =13.5 “decrease”, Hb =18 "normal", Hb=18 "hyper" | LA abnormal | LA abnormal |
| BUN               | BUN=15 “decrease”, BUN =40 “normal”, BUN=40 “hyper” | Patient Status In Follow Up 2 | Follow =follow up, Non= Death / Hospitalization |
| CR                | CR=0.7 “decrease”, CR <=1.4 “normal”, CR=1.4 “hyper” | Patient Status In Follow Up 3 | Follow =follow up, Non= Death / Hospitalization |
| NA                | NA=155 “low”, NA=148 “normal”, NA=148 “hyper” |

Table 1. Variables discretization in chronic heart failure follow up

Figure 1. Decision tree in first follow up (C4.5)
functions have been done on data for each structural integrity data. At the end, data was attached to x-validation function. Inside this function was used C4.5 decision tree for modeling. Figure1 show the Inside the x-validation function in the first follows up.

The same procedure was done on data of second follow up but this time the patient condition in third status was considered as a label. Also CHAID decision tree was used for modeling. In third phase of follow up of CHF patient, the patient condition is not specified, so gathered data created before and prediction models were evaluated. If there was an agreement between models, predicted value were considered as label and if the response was different, non-follow up label was considered because of the importance of the patient’s condition, that was admitted in hospital in one of the before phases. Also ID3 decision tree was used for modeling in this section. All these described phases of data analysis with classification methods and C4.5, CHAID, ID3 decision tree algorithms were show in Figure2.

For the K-NN model used similar structure due to the properties of the tree structure shown. K-NN model was done in three parts of follow up like decision trees models; the difference was that inside the x-validation function K-NN beside the decision algorithms were used. In these models with see the output can understand the similarity of the patient symptoms to each other. With retrieval of k-neighbor on each attribute in many cases results lead to more

only clinical and demographic factors and drug usage in order to investigate the effect of them on length of hospitalization for heart disease in general not specifically chronic heart failure had been studied. However common factors in two studies shows that studied factors, in addition to the impact of deteriorating chronic heart failure patients after discharge can affect length of stay of these patients. So, consideration of these factors in intelligent data analysis is very important.

Karalis and et al (2010) were assessment of the risk factors of coronary heart events based on data mining with decision trees. They said age, sex, family history, smoking, hypertension and diabetes are most risk factors that should be considered (17). All of the data obtained in Karalis et al study are matching with the minimum data sets for follow up and monitoring of chronic heart failure patient status in demographic sections of this research.

Rajeswari and et al (2011) with artificial neural network techniques were studied following items as risk factors for ischemic heart disease. Age, sex, cholesterol, LDL, smoking, life style, diet habits, family
Disorders and genetic factors (18). Some of these factors are similar to the results of our research. This indicates that it is important to examine these factors in follow up of chronic heart failure patients.

Dessai (2013) suggest intelligent heart disease prediction system with using probabilistic neural network. In this system 13 attributes based on k-mean clustering algorithms and artificial neural networks have been studied. This research stated that the application of intelligent system compared to the existing methods has significant impact on improving the predicting and classifying of heart diseases (19). Chen and et al (2011) suggest heart disease prediction system (HDPS). In this system also data of 13 attributes of patients were collected and use artificial neural network algorithm. Accuracy prediction in this model was 80 percent (20). In both of systems 13 attributes were studied but in this research 44 attributes in six sections studied in three different times after discharge of chronic heart failure patients. The performances of the proposed model in addition to prediction are including data processing, evaluation, reasoning, negotiation for make a decision. Also this model has learning ability and will significantly help to making appropriate decisions about the chronic heart failure patients after discharge. Follow up plan management is very essential and has significant role in reduce mortality and morbidity, decrease costs and improve use of health care resources.

Data analysis with artificial intelligent techniques is main tools that can provide access to hidden knowledge in large database of clinical and administrative data in hospitals. Also provide useful knowledge for health care personnel to improve decision support, increase quality of preventive; treatment and diagnosis planning. Nowadays data analysis without tools based on intelligent analysis is very difficult and sometimes impossible due to advances in the field of health and increasing complexity of health data. Intelligent data analysis increase efficiency of knowledge discovery process and avoids reporting trivial results to the user.

Our emphasis in this study were on non-follow up patient status include death or/and hospitalization. In data analysis system learns from its mistakes and errors lead patient to visit the physician. In other words error generated in data analysis was such that patient who had not been seen in survey their status by cardiologist lead to visit the doctor. Of course gradually and over time the number of errors will lower. All algorithms are able to predict patient statues with various degrees of accuracy. According to various aspects of data, different decision trees models used in data analysis and finally system can be voting from these decision trees. Combining decision trees based on voting algorithms can learn multiple models and predictions can reduce error of decision trees on real environment problems. Also in order to increase prediction accuracy, greater volume of data should be collected.

5. CONCLUSION

Physicians in these days rely on evidence-based medicine more than facts that generated by software. However, intelligent data analysis because of diversity of health data and specific circumstances of health domain confront enormous challenges. Limits the number of disease-related data and lack of electronic guidelines and expert system that help to physician in decision making are some challenges that this research is also facing them.

Creating and using standard terminologies and databases consistent with these terminologies help to meet the challenges related to data collection from various places and application data in intelligent data analysis.

It should be noted that intelligent analysis of health data and intelligent system can never replace with cardiologists. It can act as a helpful tool for the cardiologist’s decisions making. Lack of physician trust in intelligent analysis is the main problems of using intelligent system in the world. One of the positive points of this project is involvement of cardiologist besides the software and health information management experts in order to decrease this challenge. Explain benefit of intelligent system to cardiologist, emphasizing the helpful role of these types of systems, report generated results and feedbacks to them can decrease physician resistance and improve their attitudes and trust.

CONFLICT OF INTEREST: NONE DECLARED.

REFERENCES
1. Mohammadzadeh N, Safdari R, Rahimi A. Multi-Agent System as a New Approach to Effective Chronic Heart Failure Management: Key Considerations. Healthc Inform Res. 2013; 19(3):162-166.
2. Klersy C, De Silvestri A, Gaburri G, Raisaro A, Curri M, Regoli F, Auricchio A. Economic impact of remote patient monitoring: an integrated economic model derived from a meta-analysis of randomized controlled trials in heart failure. European Journal of Heart Failure. 2011; 13: 450-459.
3. Allen LA, Stevenson LW, Grady KL, Goldstein NE, Matlock DD, Arnold RM. Decision Making in Advanced Heart Failure: A Scientific Statement from the American Heart Association. Circulation. 2012; 125: 1928-1952.
4. Ross JS, Chen J, Lin Z, Bueno H, Curtis JP, Keenan PS, Normand ST, et al. Recent National Trends in Readmission Rates After Heart Failure Hospitalization. Circ Heart Fail. 2010; 3: 97-103.
5. Gallacher K, May CR, Montori VM, Mair FS. Understanding Patients' Experiences of Treatment Burden in Chronic Heart Failure Using Normalization Process Theory. Ann Fam Med. 2011; 9: 235-243.
6. Lainscak M, Blue L, Clark AL, Dahlstrom U, Dickstein K, Ekman I, et al. Self-care management of heart failure: practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. European Journal of Heart Failure. 2011; 13: 115-126.
7. May CR, Finch TL, Cornford J, Edley C, Gately C, Kirk S, Jenkins KN, et al. Integrating telecare for chronic disease management in the community: What needs to be done?. BMC Health Services Research. 2011; 11: 131.
8. Finkelstein J, Wood J. Delivering chronic heart failure telemanagement via multiple interactive platforms. J Syst Cybern Inf. 2013; 11(3): 34-39.
9. Lavrac N, Keravnou E, Zupan B. Book Review: Intelligent Data Analysis in Medicine and Pharmacology. Artificial Intelli-
INSTRUCTIONS FOR THE AUTHORS

ALL PAPERS NEED to be sent to: Editorial board of the journal “Acta Informatica Medica”, electronically over the web www.scopenmed.org and www.avicenapublisher.org. Every article sent to Acta Inform Med gets its number, and author(s) will be notified if their paper is accepted and what is the number of the paper. Original paper could not contain more than 3,000 words, Review article more than 4,500 and Case report more than 1,500 words, including References.

EVERY CORRESPONDENCE WILL use that number. The paper has to be typed on a standard format (A4), leaving left margins to be at least 3 cm. All materials, including tables and references, have to be typed double-spaced, so that one page has no more than 2000 alphanumerical characters (30 lines) and total number of used words must not be more than 3,500. Presenting paper depends on its content, but usually it consists of a title page, summary, text, legends, references for pictures and pictures. Type your paper in MS Word and send it on a diskette or a CD-ROM, so that the editing of your paper will be easier.

EVERY ARTICLE has to have a title page with a title of no more than 10 words: name(s), last and first of the author(s), name of the institution the author(s) belongs to, abstract with maximum of 45 letters (including space), footnote(s) with acknowledgments, name of the first author or another person with whom correspondence will be maintained. THE PAPER NEEDS to contain structured summary, 200 words at the most. Summary needs to hold title, full name(s) and surname(s) of the author(s) and coauthor(s), work institution, and all essential facts of the work, introduction, formulation of problems, purpose of work, used methods, (with specific data, if possible) and basic facts. Summary must contain the re-view of underlined data, ideas and conclusions from text. Summary must have no quoted references. Four key words, at the most, need to be placed below the text.

AUTHENTIC PAPERS CONTAIN these parts: introduction, goal, methods, results, discussion and conclusion. Introduction is brief and clear review of the problem. Methods are shown, so that interested reader is able to repeat described research. Known methods don’t need to be identified, they are cited (referencing if being listed, their generic name is used, (brand name can be written in brackets). Results need to be shown clearly and logically, and their significance must be proven by statistical analysis. In discussion, results are interpreted and compared to the existing and previously published findings in the same field. Conclusions have to give an answer to author’s goals. QUOTING REFERENCES MUST be on a scale, in which they are really used. Quoting most recent literature is recommended. Only published articles, (articles accepted for publishing), can be used as references. Not published observations and personal notifications need to be in text in brackets. Showing references must be as how they appear in the text. References cited in tables or pictures are also numbered according to the quoting order. All references should be compiled at the end of the article in the Vancouver style or pubMed style (i.e. www.scopenmed.org).

TESTS USED for statistical analysis need to be shown in text and in tables or pictures containing statistical analysis. TABLES HAVE to be numbered and shown by their order, so they can be understand without having to read the paper. Every column needs to have a title, every measuring unit (SI) has to be clearly marked (i.e. preferably in footnotes below the table, in Arabic numbers or symbols).

Pictures also have to be numbered as they appear in the text, drawings need to be enclosed on a white or tracing paper, while black and white photos have to be printed on a radiant paper. Legends (e.g. next to pictures and photos), have to be written on a separate A4 format paper. All illustrations, (pictures, drawings, diagrams) have to be original, and on their backs contain illustration number, first author’s last name, abbreviated title of the paper and picture at the top. It is appreciated, if author marks the place for the table or picture. Papers could not be submitted with included more of 4 Tables, Figures or Graphs. Every additional must be paid 20 euros each.

USE OF ABBREVIATIONS have to be reduced to a minimum. Abbreviations can be used without their definitions. Supplement. If paper contains original contribution to a statistical method or author believes, without quoting original computer program, that paper’s value will be reduced. Editorial staff will consider possibility of publishing mathematistics/statistical analysis in extension.

ANY PRACTICE OF plagiarism will not be tolerated regarding submitted articles. Non-identifiable quoted parts of the articles from other authors are known act of plagiarism if it is not cited or referencing in appropriate places in the article. Advertent practice of plagiarism will abort reviewing process or article submission. Author(s) may suggest or exclude peer-re-viewers for their articles but Editorial Board has the right to reject their(s) opinions or suggestions according to copyright Assignment form signed by authors before reviewing process. Authors must respect guidelines and rules of ICMJE, WAME, COPE, E A SE, linked on www.avicena-publisher.org.

ALL INDIVIDUALS LISTED as authors should qualify for authorship and should have participated sufficiently in the work to take public responsibility for appropriate portions of the content and follow the next conditions: a) substantial contributions to the conceptions and design, acquisition of data, or analysis and interpretation of data; b) drafting the article or revising it critically for important intellectual content; c) final approval of the version to be published (all authors must sign copyright Assignment form downloaded from www.avicena-publisher.org). All other contributors to the article’s subject who does not qualify for authorship should be acknowledged in acknowledgement section, for all relevant information about authorship follow ICMJE guidelines.

ALL AUTHORS MUST make a formal statement at the time of submission indicating any potential conflict of interest that may constitute an embarrassment to any of the authors if it were not to be declared and were to emerge after publication. Such conflict of interest might include, but not limited to, share holding in or receipt of grant or consultancy fee from a company whose product features in the submitted manuscript or which manufactures a competing product. All authors must submit a statement of conflict of interest to be published at the end of their article (conflict of Interest: non declared).