International Journal of Research in Science and Technology
Volume 6, Issue 3: July - September, 2019
ISSN 2394 - 9554

CLASSIFICATION OF ICD-10 CODES FOR DIAGNOSTIC AND MONITORING OF HEALTHCARE DEVICES

Raj Kumar
Assam Engineering College, Guwahati, Assam, India

ABSTRACT
The World Health Organization (WHO) distributed a 10th revision of the International Statistical Classification of Dis- eases and Related Health Problems (ICD-10) . ICD-10 contains data about diseases, devices used to cure the diseases, external causes of injuries and diseases, ICD-10 codes, and other related information. This research study addresses the classification of ICD-10 codes based on medical devices associated with adverse incidents in diagnostic. Decision tree grouping method will be utilized in this research study to distinguish the connection between the mortality rate and the causes of the mortality. The findings of this research study will help the hospitals and the healthcare devices service providers to understand how effectively they can manage the healthcare device maintenance to provide better patient care. According to the best of my knowledge, this is the first research on mortality rate analysis based on improper use of healthcare instruments/devices using ICD-10 codes.

Keywords: World Health Organization, healthcare, Health Problems

I. INTRODUCTION
The healthcare industry is growing at an alarming pace and it is interesting to see how data mining can help to achieve the goal of better service to patients in need. Analyzing mortality data is important to understand the complex circumstances of death across the country. Since healthcare systems follows the International Classification of Diseases (ICD-10), it is easy to classify patient deaths based on the causes. We will study the mortality data set and analyze the number of deaths that occurred due to a defective healthcare instrument/device used in the treatment of the diseases. Every year, the Center for Disease Control and Prevention (CDC) releases a detailed report on the deaths in the United States with causes and related ICD-10 codes. In this research study we will analyze the mortality rate based on the causes related to defective healthcare instruments used for the treatment of the diseases. The research study will use a simple decision making classification technique to reduce the complexity of the system. The results of this research study will help the healthcare instrument providers to monitor and provide proper diagnostics for the instruments. This will ultimately reduce the deaths because of using defective healthcare instruments in the treatment of the patients. Though the code related systems have been used for a long time, future study will include to study the challenges of implementing ICD-10 codes. As the ICD codes are updated by the WHO with the advancement of technologies, the future work of this research work will include advancing the classification techniques accordingly.

II. BACKGROUND
A. Subject Area
The subject area for our research study is the diagnostics of healthcare devices based on the classification of death causes. The migration of the coding system ICD-10 by the World Health Organization (WHO) allows healthcare providers to categorize diseases and track healthcare and medical compli- cations effectively. The migration has increased the number of codes from 14, 000 to 17, 000 that has helped the health- care industry to seize more granular data. These codes and their fundamental value of real-time, provider-based and point-of-care coded assessment of diseases is important as they provide valuable insights to patient care. There has been various studies in the field of healthcare by using the ICD-10 coding system. The research paper presents how ICD-10 can be used to predict in-hospital mortality. describes the ICD-10 classification of mental and behavioral disorders. The study of ICD-10 codes will be helpful in reducing the improper use of healthcare instruments and devices.

B. Approach
ICD-10 codes offer detailed and increased ability to ac- commodate new findings and technologies. These codes can help evaluate and improve the quality of patient care if used properly. The classification technique can be very useful to identify patients for potential diseases. Classification methods can additionally be used to predict the cost of treatment for medical services. Classification technique is a data-mining technique that organizes the related entities into one set. The study will help us discover cost savings for patients and

improved monitoring of patients. As the data inside ICD-10 is already classified, it is easy to apply a classification technique like the decision tree and find solutions to the improper use of the healthcare devices.

C. Key Concepts

The key concepts while conducting any research related to ICD-10 are: 1) ICD-10 is far more detailed than ICD-9, with 8,000 categories compared to 5,000 categories 2) ICD - 10 uses alphanumeric codes 3) The ICD-10 codes are divided based on the various diagnosis and procedures the doctors provide.

III. OBJECTIVES

To classify the data into high risk, low risk, and medium risk classes, the data need to be cleansed and profiled in a way that suits the data-classification algorithm. It is important to understand the qualifiers that will be used to solve the problem statement. Diagnostic and monitoring healthcare de- vices used for anesthesiology, cardiovascular, neurological, or other purposes will help us understand and classify the major problems in healthcare due to device associated incidents. The classification of these device-related incidents will help us understand and classify the patient's death potential risk as high, medium and low. In addition to classification, this will help the hospitals in spending some more time to do analysis and help them decide if a new device is needed for diagnosis.

IV. METHODS

There are various methods for solving classification problem. We will use decision tree as a classification method and the classification algorithm described in Figure 1. The decision tree is identical to the flowchart. Each non-leaf node denotes a check on a particular attribute. Every branch denotes a result of that check and each leaf node has a class label. A tree like graph is used in the decision tree classifier. Using a decision tree, decision-makers will opt for the best alternative. Traversal from the root to a leaf shows distinct class separation based on most statistics gain. The decision tree approach is used frequently with the helpful resource of the many researchers within the health care field. In our research study, we are going to develop a small decision tree. If the decision tree is small, the cost required to set up the decision tree will also be less. If the decision tree is the bigger one, the cost required to set up the tree is more and the prediction is also poorer.

A. Data Collection and Profiling

The data will be collected for various ICD codes related to the device diagnostic and monitoring for various procedures. The ICD-10 data will be collected from the WHO website by data-scrubbing. Python will allow to scrub and cleanse the data in the required format. The mortality data set is available on the Kaggle website. The data set available is present in **.pkl, CSV, and JSON formats.**

B. Classification Algorithm

The classification type applied in 1 can be addressed as follows. The training set (can be also referred to as an input data), includes more than one record, having more than one attribute. All the attributes are given an uncommon class name. The goal of the classification technique used in this research study is to study analyze the training data set and to develop an accurate model or a description for each class using the data features. If the class labels are unknown for the test data, the class descriptions can be used. They can additionally be used to develop a higher grasping of every class in the data . Our goal is to classify large data sets, hence this research study will focus on decision tree classifiers.The reason for this is that decision tree classifiers are significantly fast compared to alternative classification methods. There are other methods like neural networks, but they will have very prolonged training time even for small information sets. The decision trees are often remodeled into simple and convenient to grasp classification rules.

 ISSN 2394 - 9554

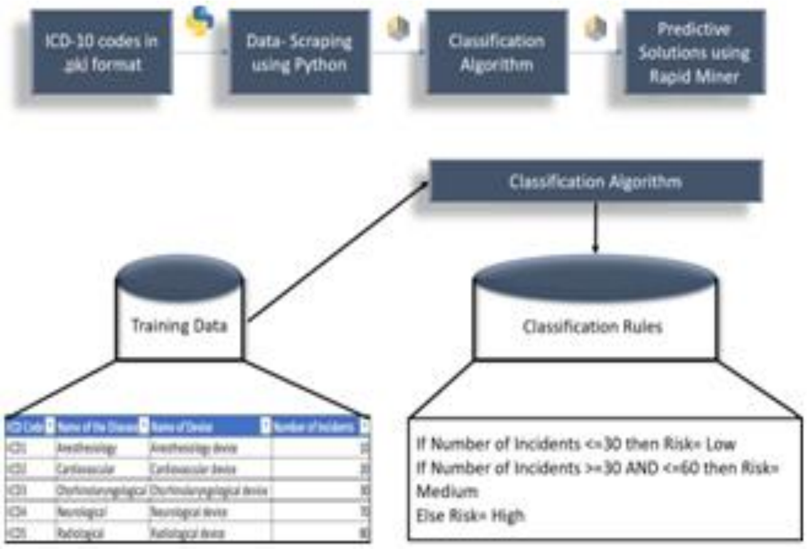

Fig. 1. Classification Algorithm

C. Analysis

Once the data is collected and the classification rules are applied, it is important to analyze the outputs. The outputs will help us know the number of patients affected classified by the risk. We can check the analysis by calculating how many times the ICD code is repeated in the data set. The higher number of times the ICD code is repeated, the higher is the risk of patient death.

D. Prediction

Furthermore, analysis and the application of business in- telligence will help the hospitals predict if newer diagnostic instruments are needed and to understand the cause of death due to medical device failure. This prediction will be helpful to understand in detail the life wear and tear of the diagnostic devices. The *Rapidminer* tool will be used for building pre- dictive solutions. Rapidminer is a user-friendly software for data mining strategies and machine learning. RapidMiner provides an advanced analytical solution through template- based frameworks. These template-based frameworks help to reduce errors. One more advantage of using Rapidminer is that it eliminates the need to write a code. It represents a standard approach to design even very complicated issues . Rapidminer has versatile operators for information input and output in several file formats. It contains various learning schemes for classification, clustering and clustering tasks.

E. Data set

The data for diagnostic and monitoring is made available on the Kaggle website. The mortality data set we are using is a record of every death in the *United State*s for 2005 through 2015, including detailed information about the causes of death and the demographic background of the deceased. For data scraping from ICD codes, Python programming language will be used as the data is available in .pkl file formats. Once the data is extracted the data can be cleansed using Python or any other open-source tool. The data set will need some profiling and cleansing to pass it to the downstream classification algorithm. The classification algorithm will play a major role in the analysis. The classification algorithm will classify the associated risks in buckets and help in further analysis and prediction. The mortality data set will include attributes [17] as follows: resident status, sex, age, marital status, autopsy, ICD-10 code, race. Each data set is a collection of one year of data which is paired with JSON files and an ICD-10 code set.

V. EXPECTED RESULTS

The focus of the research study is classifying the mortality rate based on the ICD-10 codes. A root cause analysis will be done to maintain the accuracy of the research study. It is expected that the research will produce results without any fail because the data is processed in different stages. The data processed will be profiled in a way that the reporting tool can build bar-charts, pie-charts, and predictive dashboards easily. The results will give an insight into how many deaths are being caused due to the improper use or use of faulty medical instruments in the treatment. The classification rules mentioned in this paper will be used to categorize the death incidents. If the number of incidences are less than 30, they will be classified as low risk. If the number of incidences are between 30 and 60, they will be classified as medium risk. If the number of incidences are more than 60, then it will be classified as high risk. The ICD-10 code data will be mapped with the mortality rate data. The size of the dataset will depend on the number of deaths related to the inaccurate/faulty medical instruments. The results of this research study will be displayed using Rapidminer, to make them easy to understand and visualize. The data visualization will be presented in a Business Intelligence format which will give more insights regarding the data as well as the causes. The final dataset will contain the following attributes: ICD-10 code, death cause, name of the instrument used in the treatment of the patient, name of the disease, number of the incidents. The second dataset will have the information about the low risk, medium risk and high-risk incidences as mentioned above.

VI. CONCLUSION AND FUTURE WORK

In this research study, we have proposed a classification algorithm which will be useful in classifying mortality rate using ICD-10 codes. To classify the mortality risks we are using a small decision tree to avoid complexity and produce accurate outputs. It is important to understand how a simple

classification algorithm can be used to predict mortality rates. This research study also serves a baseline for classifying various reasons for mortality rate other than the use of faulty instruments. In addition to the classification of the risks associated with the number of incidents, these research can also be used as a prototype for predicting the life of the medical instrument. In addition to this, the results can serve as a reference to the hospitals and the medical instrument service providers to effectively manage the maintenance of these instruments.

REFERENCES

1. M. Minin˜o, R. N. Anderson, M. A. Boudreault, and M. Warner, “Deaths: Injuries, 2002,” National Vital Statistics Reports, vol. 54, no. 10, 2006.
2. Naik and L. Samant, “Correlation review of classification algorithm using data mining tool: Weka, rapidminer, tanagra, orange and knime,” Procedia Computer Science, vol. 85, pp. 662–668, 2016.
3. J. Cartwright, “Icd-9-cm to icd-10-cm codes: what? why? how?” 2013.
4. Tomar and S. Agarwal, “A survey on data mining approaches for healthcare,” International Journal of Bio-Science and Bio-Technology, vol. 5, no. 5, pp. 241–266, 2013.
5. Van Rossum and F. L. Drake, The python language reference manual. Network Theory Ltd., 2011.
6. M. G. Weiner, “Point: Is icd-10 diagnosis coding important in the era of big data? yes,” Chest, vol. 153, no. 5, pp. 1093–1095, 2018.
7. M. Kamber, L. Winstone, W. Gong, S. Cheng, and J. Han, “Gen- eralization and decision tree induction: efficient classification in data mining,” in Proceedings Seventh International Workshop on Research Issues in Data Engineering. High Performance Database Management for Large-Scale Applications. IEEE, 1997, pp. 111–120.
8. M. Mehta, R. Agrawal, and J. Rissanen, “Sliq: A fast scalable classifier for data mining,” in International conference on extending database technology. Springer, 1996, pp. 18–32.
9. R. Escorpizo, N. Kostanjsek, C. Kennedy, M. M. R. Nicol, G. Stucki, and T. B. U¨ stu¨n, “Harmonizing who’s international classification of diseases (icd) and international classification of functioning, disability and health (icf): importance and methods to link disease and functioning,” BMC Public Health, vol. 13, no. 1, p. 742, 2013.
10. R. Khera, K. B. Dorsey, and H. M. Krumholz, “Transition to the icd-10 in the united states: an emerging data chasm,” Jama, vol. 320, no. 2, pp. 133–134, 2018.
11. S. E. Bowman, “Why icd-10 is worth the trouble,” Journal of AHIMA, vol. 79, no. 3, pp. 24–29, 2008.
12. T. B. Sanders, F. M. Bowens, W. Pierce, B. Stasher-Booker, E. Q. Thompson, and W. A. Jones, “The road to icd-10-cm/pcs implementa- tion: forecasting the transition for providers, payers, and other healthcare organizations,” Perspectives in health information management/AHIMA, American Health Information Management Association, vol. 9, no. Winter, 2012.
13. T. Y. Lin, “Attribute (feature) completion-the theory of attributes from data mining prospect,” in 2002 IEEE International Conference on Data Mining, 2002. Proceedings. IEEE, 2002, pp. 282–289.

14. V. Sundararajan, T. Henderson, C. Perry, A. Muggivan, H. Quan, and W. A. Ghali, "New icd-10 version of the charlson comorbidity index predicted in-hospital mortality," Journal of clinical epidemiology, vol. 57, no. 12, pp. 1288–1294, 2004.
15. W. H. Organization et al., The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1992.
16. W. H. Organization, International statistical classification of diseases and related health problems: instruction manual. World Health Organization, 2004, vol. 2.
17. Z. Yao, P. Liu, L. Lei, and J. Yin, "R-c4. 5 decision tree model and its applications to health care dataset," in Proceedings of ICSSSM'05. 2005 International Conference on Services Systems and Services Management, 2005., vol. 2. IEEE, 2005, pp. 1099–1103.