Research Article

iMedPub Journals
www.imedpub.com

American Journal of Computer Science and Information Technology

2020
Vol.8 No.2:53

DOI: DOI: 10.36648/2349-3917.8.2.53

Lower Respiratory Tract Infection Clinical Diagnostic System Driven by Reduced Error Pruning Tree (REP Tree)

Olayemi Olufunke Catherine*

Department of Computer Science, Joseph Ayo Babalola University, Ikeji Arakeji, Osun State, Nigeria

***Corresponding author:** Olayemi Olufunke Catherine, Department of Computer Science, Joseph Ayo Babalola University, Ikeji Arakeji, Osun State, Nigeria, E-mail: ocolayemi@jabu.edu.ng

Received date: June 13, 2020; **Accepted date:** June 29, 2020; **Published date:** July 06, 2020

Citation: Catherine OO (2020)Lower Respiratory Tract Infection Clinical Diagnostic System Driven by Reduced Error Pruning Tree (REP Tree). Am J Compt Sci Inform Technol Vol.8 No.2: 53

Abstract

Combating Lower Respiratory Tract Infections (LRTIs) in Peadiatrics has prompted the development of this new diagnostic model. It was confirmed from several literature reviews that the Lower respiratory tract infections accounts for over a million children illness and death yearly as the effect of lack of prompt diagnosis, or no diagnosis due to shortages of medical experts and medical facilities in our localities. This new diagnostic model was built by applying Reduced Error Pruning Tree (REP Tree) Algorithm on the LRTIs data sets collected from a Federal Medical Center Owo, in Ondo State. When the model was tested, it presented 100% detection proportion on the training cases and 95.7142% success proportion on the testing cases. It is sure that the full implementation of this presented model (rules generated from the REP Tree) on any platform will decrease the high death rate associated with respiratory infections in peadiatrics.

Keywords: Cyanosis; lower respiratory tract infections; Paediatric; Machine learning; REP Tree

Introduction

Data mining is a procedure that is being used by researchers to transform raw data into valuable Information. One of the primary goal of data mining is to predict an unknown value of a new sample from observed samples, such as a prediction is achieved by two sequential phases. The first phase is the training phase which produces a prognostic model from training samples which make use of the available supervised learning techniques. The second phase is the testing phase used for evaluating the general predictive model that uses the testing samples that has not been used in the first stage which is the training stage [1].

Health Specialists are not always accessible in most of the communities today. Even when available, it is expensive that many parents cannot afford the payment. World Health Organization (WHO) in their report of 2012 reports that the population of health staffs are not increasing as the global population is increasing. The few available health officers are based in the big cities, while the rural areas are lacking [2]. Medical Knowledge is speedily increasing today and this expansion is making computer aided diagnostic system desirable. In the recent development of Artificial Intelligence (AI), this present technique as the prospect of solving tasks that are initially complex to solve with computer-based systems in the medical domain [3]. Since it was reported by WHO that health specialists are scarce, then one of the method that can be used to reduce this deficiency of medical experts is by developing an e-health which aims at providing health services through information system medium. These ranges from tele-health system which is the presentation of health services and information via telecommunication technology to specialized expert system developed to perform the duty of expert in a specific health care [4]. Diagnosis of transmittable sicknesses is a work that enables doctors to make prognosis about features of medical conditions and to determine suitable solutions to it [1]. Generally, this research is focusing on the new applications in the health field and particularly diagnosis of different diseases [5]. Computer skills has been effectively applied in medical field over the years to implement diagnosis and treatment in the form of medical decision support systems and this method is fast growing day by day in different areas of medical problems. An accurate medical diagnosis will surely ensure proper treatment of the diagnosed disease or illness. LRTIs are the main reasons of death in the developed and developing countries. Respiratory Infections (RI) are accountable for over 5 million deaths per year. In the industrialized countries, it is also a major reason of outpatient consultations. The human Respiratory tract is divided into the Upper and the Lower Tracts. The upper RTIs are located in the respiratory tract situated above the vocal cords and without signs on auscultation, LRTIs includes the entire conditions which might or may not involve the parenchyma. The parenchyma is the vital tissue as distinguished from the connective and supporting tissue in the human body. The infections that is involves the parenchyma is identified as pneumonia, also, the Infections that are not connected to the parenchyma are called acute bronchitis in paediatrics patients [6]. Respiratory diseases among children are major reasons for death and sicknesses worldwide especially among children under the ages of 5 years in the developing Countries [7] .On the average, it has been reported that every child has about 5 to 6 episodes of LRTIs in a year, accounting for about 30 %-50% of the total paediatrics outpatient visits Data from National Demographic Health Survey 2013 reported the prevalence of acute respiratory infections (ARI) in Nigeria to be about 20% [8,9]. Chronic LRTIs diseases rank as the third leading cause of death in the United States [7]. Generally, over 2.5 million deaths occur annually from these respiratory tract infections and these contribute to the environment due to these infections. It is the second leading killer of young children around the world due to the attitude of care givers, causing approximately 20% of all children's deaths [10].

Related Literature

Several researchers in Medical Informatics domain has worked in the area of transmittable sicknesses such as LRTIs so as to decrease the figures of deaths linked with it. Some researchers in the domain of health have developed some computer aided diagnostic and treatment systems for LRTIs, where they all emphasized the need for further research, some of these works are described below:

Oguntimilehin et al. [4] presented a paper with the aim classifying patients having respiratory infections among some set of children admitted in a clinic using machine learning technique. Inadequate number of medical specialists which has increased the mortality of patients who suffer from LRTIs and the need to use computer technology to reduce the number of mortality and the waiting time to see the specialist on infections in paediatrics necessitated this study. Inability to evaluate of the symptoms of the infections and as well as to evaluate the degree of the illness are the major weaknesses of this system. Some authors in worked on a paper work due to the reason being that most of the existing systems on Respiratory diagnosis failed to provide therapy while some provide therapy without diagnosis [11]. Also, because half of the world's population is at the risk of infections and deaths associated with Respiratory Infection are at increasing rate. A machine learning Technique Naive Bayes was used to generate a classification model for RIs diagnosis for different respiratory infection cases and therapy was provided accordingly. The model was derived from a small data set (120 data samples). A larger sample will reveal a better diagnostic pattern

Adiku et al. presented Acute LRTIs in children under the age of five years in Accra Ghana. The aim of the work was to examine the aetiological agents and symptoms that are related with ALRI in children below the age of five years old at the Kurdebu teaching hospital [7]. Nasopharyngeal aspirates and venoms blood specimen were obtained from 108 children with features suggestive of ALRI were cultured and isolated bacterial organisms were also identified biochemically. Social-demographic and clinical data were also obtained from the study subjects. The methods used in diagnosing the diseases were not specified in this paper. The limitation of the study is that, ALRI surveillance did not cover a full year. There were no algorithms used in this study, limited dataset and limited attributes were used in the study.

The Epidemiology of LRTIs in children was presented in to offer clinicians a brief update on the recent epidemiology of RIs in peadiatrics. They presented significant risk factors as a role in the cases of respiratory infections in children. They identified the most significant variables effects as parental smoking, zinc deficiency, mother's experience as a caregiver and malnutrition. The methods used for the study and implementation were not mentioned.

The risk factors of childhood acute LRTIs in Northern Nigeria as presented in was motivated by the need to identify risk factors of ALRTIs among children below the ages of five years that were hospitalized [12,13]. A prospective and descriptive study of children between the ages of 2- 60 months was carried out in which the descriptive statistical analysis (chi-square) was used to calculate the minimum sample size [14]. The result shows that younger age under 24 months and exposure to hydro-carbon and biomas from indoor pollution were contributing risk factors for ALRTI. Cyanosis was not considered as a variable and also there was no feature selection method used in the study so as to determine the most relevant variable as a risk factor for the infection [15-17]. The size of the dataset used is too small to ascertain the veracity of the results presented.

Materials and Methods

Research, review and medical consultation

The Analysis of some available medical support systems on diagnosis and treatment of LRTIs in the field of medical informatics was carried out with the intension of improving on their weaknesses so as to have a promising new diagnostic system for this type of infections. Several health specialists were consulted for the success of this system.

Data collection and description of data sets

Data on LRTIs cases diagnosed through clinical diagnosis method were collected for a period of three months from FMC Owo, Ondo state, Nigeria. Seven hundred and two (702) patients records were collected having twenty three (23) attributes with Class ID of which four hundred and sixty (460) were infected and two hundred and forty (240) were non infected. The features in the dataset are listed as follows; {Sex, Age, Weight, Nutritional Status, Breast feeding, Parental smoking, Cyanosis, Respiratory Rate, Cough, Temperature, Low birth weight, Indoor air pollution exposure, Incomplete immunization, Crowding, Under- nutrition, Fever, HIV infection, Difficulty in breathing, Heart rate (pulse), Educational Status, Daycare, Herbal Mixture and Class Id}. The records were splited into two datasets such as training and testing, 70% and 30% respectively. To avoid bias, the records for each set were selected randomly The data collected was processed for the presence of error in data entry including misspellings and missing data. The data stored in the comma separated variable (.csv) format was transformed into the attribute file format (.arff) so as to develop predictive model for the diagnosis of LRTIs for the peadiatrics patients.

The REP tree

Reduced Error Pruning Tree (REP Tree) is a Machine Leaning technique and a fast decision tree learner

$$info(D) = -\sum_{i=1}^{m} p_i \, log_2(p_i) ---------(1)$$

This technique uses the information gain given in Equation 1 to determine the splitting Node (N) which represents the tuples of partition D, where This technique uses the information gain given in Equation 1 to determine the splitting Node (N) which represents the tuples of partition D, where i=probability that an arbitrary tuple in D belongs to Class i and is estimated by |i, | 1||. It uses Reduced Error Pruning (REP) method with back fitting for pruning and a tree is generated not direct rules. It uses first-better search strategy and a post order traversal for searching in the pruning space. The evaluation function is defined in Equation 2.

$$f(T) = -\sum_{t \in Y_r} e(t) -------------(2)$$

where e(t) is the number of errors made by node t during the classification of the examples in the pruning set. The search in the space moves from a state.

1€r (T) if the inequality (′) ≥ () holds using bottom up approach or equivalently if

$$\sum_{t \in y_r} e(t) \leq \sum_{t \in y_r} e(t) ----------(3)$$

The reason is to estimate each non- terminal node t regarding the classification error in the pruning set. If this error decreases subtree ′ rooted on t is replaced by a leaf node, then ′ must be pruned. The REP Tree Algorithm was used in building a classification model in the form of a Decision Tree for Lower Respiratory Tract Infection Diagnosis.

Experimental Set Up, Results and Discussion

The values of the Identity Class used in this paper work were converted to Integer value as follows: Infected (Yes)=1 and Non-Infected (No)=0 and the decision attribute classes were used as follows: High=3, Mod=2 and Low=1.

Rep Tree algorithm has described above was used on the seven hundred and two (702) instances ,70% used for training and 30% for testing to build a classification model for LRTIs diagnosis in form of a Decision Tree. The Decision Tree generated from the REP Tree is displayed in **Figure 1.** The model was tested on both the training set and the testing set. The confusion Matrices of the results are displayed in **Tables 1 and 2**.

Table 1: Confusion Matrix of the LRTIs diagnosis model on the Training Set.

Predicted as Actual	High	Mod	V. Low
High	194	0	0
Mod	0	488	0
V.low	0	0	20

Note Mod means Moderate and V. Low means Very Low

TP=Class group correctly classified TN=Class group incorrectly classified

$$\text{Detection Rate=} \quad \frac{TP}{TP+T}$$

$$\frac{194+488+20}{194+488+2} = \frac{702}{702} = 100\%$$

Table 2: Confusion matrix of the LRTIs diagnosis model on the training set.

Predicted as Actual	High	Mod	V. Low
High	33	0	0
Mod	0	153	0
V.low	0	0	15

The outcome of the dataset indicated that all the seven hundred and two (702) training instances were correctly classified by the LRTIs model, attaining 100% accuracy, while the two hundred and one (201) of the two hundred and ten testing instances were correctly classified attaining 95.7142% detection rate in this case. This results shows and excellent result.

Implementation

The Decision Tree generated from REP Tree will be changed to rules and the rules will be implemented as a mobile application in order to give an enhanced accessibility and wider coverage taking the advantage of the fast growing internet technology and increasing internet enabled mobile phones.

Conclusion

There is urgent need for collaboration between medical experts and Information Technology (IT) experts which has again been demonstrated in this paper work. A new diagnostic model to reduce the number of deaths and economic backwardness being caused by respiratory infections was thus developed due to synergy between IT experts and medical experts. This type of a system is needed in health care delivery so as to move the sector forward and save more lives.

Acknowledgements

The Chief Medical Director of Federal Medical Center (FMC) Owo, Ondo State Nigeria, and other medical practitioners who contributed to the success of this work are highly appreciated. Thanks to all authors whose works has been reviewed in this work.

References

1. Adetunmbi AO, Oguntimilehin A, Falaki SO (2012) Web-Based Medical Assistant System for Malaria Diagnosis and Therapy. GESJ: Computer Science and Telecommunications 1: 42-53.
2. National Population Commission (Nigeria) and ICF International (2013) Nigeria Demographic and Health Survey 2013. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International.
3. Adewumi MT, Adekunle YA (2013) Clinical Decision Support System for Diagnosis of Pneumonia In Children. Int J Adv Res Comput Sci Software Engineer 3: 40-43.
4. Oguntimilehin A, Adetunmbi AO, Abiola OBA (2015) Review of Predictive Models on Diagnosis and Treatment of Malaria Fever. Int J Comput Sci and Mobile Computing 4: 1087-1093.
5. Olayemi OC, Olasehinde OO, Ojokoh BA (2017) Development of a Predictive Model for Paediatric Patients with Lower Respiratory Tract Infection Using Bayesian Rule Approach. JABU J Sci Technol 3: 92-101.
6. Chang RBL, Rezajafari, Callanghan TO (2013) Leveraging Information Technology to Bridge the health workforce gap.Ball World Health Organ 91: 890 892.
7. Adiku TK, Asmah RH, Rodrigues O, Goka B, Obodai E, et al. (2015)Aetiology of acute lower respiratory infections among children under five years in Accra, Ghana. Pathogens 4: 22-33.
8. Achary AD, Prasanna KS, Nail S (2003) Acute Respiratory Infections in Children: Community Based Longitudinal Study in South India. Indian J Public Health 47: 1-13.
9. Mohammad RB, Ferial, Farahnaz A, Ali C, Soheila K , et al . (2013) Epidemiology of Lower Respiratory Tract Infections in Children. J Compr Ped 4: 93-98.
10. The Boston Consulting Group (2012) mHealth: An idea whose time has come. Teleomclients.
11. Dabi A, Wilson B (2003) Online decision support gives plans an edge. Managed Healthcare Executive 13: 20–24.
12. Oguntimilehin A, Adetunmbi AO, Abiola OBA (2015) Review of predictive models on diagnosis and treatment of malaria fever. Int J Comput Sci and Mobile Computing 4: 1087-1093.
13. Esposito F, Donato M, Glovanni S, Valentina T (1999)The Effects of Pruning Methods on the Predictive Accuracy of Induced Decision Trees. Applied Stochastic Models in Business and Industry 15: 277-299.
14. Ahmed PA , Yusuf KK, Dawodu (2015) A Childhood Acute Lower Respiratory Tract Infections in Northern Nigeria At Risk Factors.
15. Obot OU, Uzoka FME (2008) Fuzzy-Rule Based Framework for the Management of Tropical Disease. Int J Medical Engineering and Informatics 1:7-17.
16. Wardlaw T, You D, Newby H, Anthony D, Chopra M (2013)Child survival: A message of hope but a call for renewed commitment in UNICEF report. Reproductive Health 10:64.
17. World Health Organization Preventing a Leading Killer (2012).