IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Machine-learning Techniques for Clinical Decision-making and Prediction: A Review

K. Jeberson[1*], M. Kumar[2], R. Yadav3

[1*]Dept. of Comp. Sci. & I.T, SIET, SHUATS, Allahabad, India
[2] Indian Institute of Information Technology, Allahabad, India
[3*]Dept. of Comp. Sci. & I.T, SIET, SHUATS, Allahabad, India

[*]*Corresponding Author: klinsega.jeberson@shiats.edu.in, Tel.: +91-9451593916*

Available online at: www.ijcseonline.org

Accepted: 15/May/2018, Published: 31/May/2018

***Abstract*—** The incredible growth in medical technologies has increased in accumulation of loads of data in various forms. Application of medical informatics techniques and tools transform data into various forms of stuff which are sutable for mining. Implementing data mining techniques on clinical data enable the discovery of priceless knowledge from the huge collection of information stored. This study aims to conduct a review systematically on the classification techniques applied on clinical data from the perspective of (i) Medicine and (ii) Health care. The outcomes of this study indicate that maximum amount of research was published in the years 2015 and 2016. In medical data mining research, the most popular algorithm used was decision tree. Elsevier was identified as the leading publisher, which has published plenty of articles in this domain. 75% of the articles belonged to the category 'medicine' and rest of the articles belonged to the category 'health care'. Out of the 75% articles, most of them were related to prognosis and diagnosis of diseases and fewer studies have been conducted in treatment recommendation. Choosing the best therapy and identifying the ideal treatment plan is a challenging task in case of diseases like heart failure and cancer. Moreover there is insufficient machine-learning research conducted in kidney diseases especially in chronic kidney disease and end-stage renal disease which are considered a global threat nowadays.

***Keywords*—** Classification, Data mining, Decision tree

I. INTRODUCTION

Data mining is an prominent field and the techniques and tools have been successfully employed in solving problems in the domain areas such as banking, entertainment, health care, education and so on. In healthcare industry data mining is a used [42]. The wide ranges of applications include weather forecasting, stock prediction, medical decision support modelling and frequent item set mining. In medical field huge volumes of data are being accumulated every day from various instruments like mass spectrometry, digital microscope, MRI and so on. The scope of data mining and machine learning applications on clinical data is vast. Out of the diverse data mining techniques, predictive modelling has been immensely implemented in mining medical data. Predictive modelling has been effectively applied in various sub-domains of clinical data such as diagnosis, therapy and health care management. Studies showing the successful application of classification techniques in diagnosing diseases in various domains like oncology, endocrinology, nephrology and pulmonary diseases have been published.

In this paper, a review of the applications of data mining and machine learning techniques and tools in clinical decision making is presented. The scope of this paper is limited to classification of structured clinical data but excludes time series data. The review is structured as follows: Section II provides an introduction of data mining and machine learning, Section III provides the details of review methodology, Section IV presents an insight of literatures reviewed in "medicine" and "health care" domains, Section V provides the discussion and Section VI provides the conclusion.

II. DATA MINING AND MACHINE LEARNING

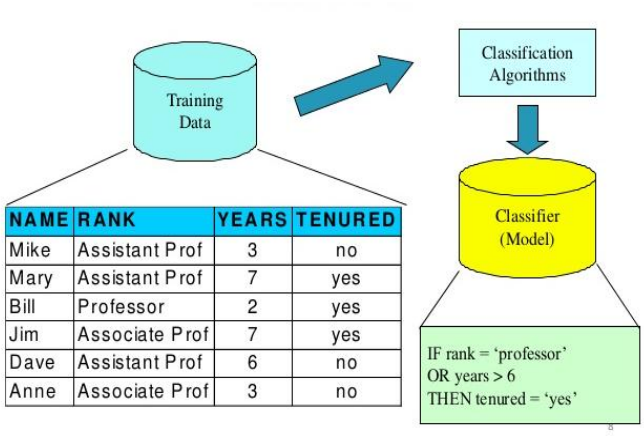
Using data mining, it is possible to mine and extract unknown knowledge from massive collection of data. The knowledge mined are stated as patterns or models. Examples of patterns include clusters, classification rules, linear models, graphs and trends in time series. The data used in mining refers to "observational data" and not "experimental data". In other words the data used in mining has been

already maintained for some purpose. For example, the transaction records maintained in a bank and medical records maintained in a hospital. This means that data collection does not form a part of data mining objectives. To this end data mining varies from statistics, where data collection is performed using efficient techniques to answer exact questions. Data mining tasks include clustering, classification, associations mining, outlier analysis and regression.

Machine learning is a subset of data mining that encompasses "supervised" and "unsupervised" learning techniques. Supervised learning use the class labels alongside other data to create models that classifies an instance in one of the classes. Classification employs supervised learning technique. In unsupervised learning the algorithms do not have any prior information of groups but create clusters based on the distance among the objects in the dataset. Clustering employs unsupervised learning technique.

A. Classification

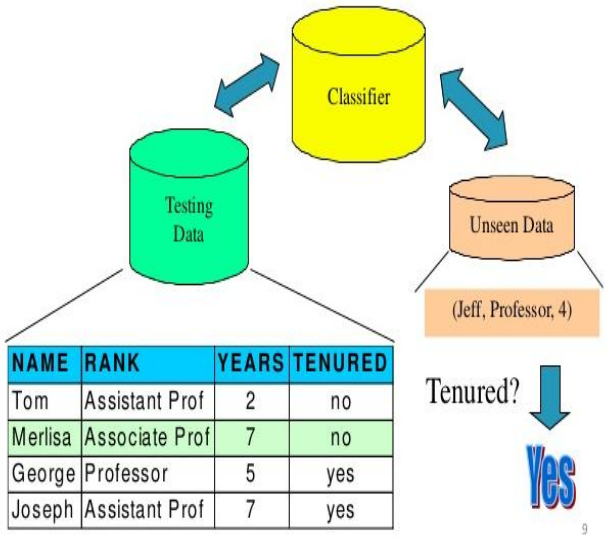
Classification is a technique in data mining, that develops models describing the various data classes. The models extracted called classifiers can be put to use to predict categorical class labels for unseen data. Classification consists of two steps: 1. Model construction and 2. Prediction. In the model construction phase also called training phase, the learning algorithm constructs the classifier from the historical data called training data (Figure. 1).



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

Figure 1. Construction of Classifier[43]

The classifier constructed can take the form of rules, decision tree, functions and so on depending on the learning algorithm used. In the prediction phase, the model is used to assign class label to the new instance (Figure. 2). An instance X with an n-dimensional attribute vector, X=(x1,x2, …xn) where x1,x2,…xn represent the n measurements of X corresponding to the n attributes A1, A2, A3, …,An of the database. The instance X is assigned a class label by the classifier. The class attribute is categorical in which the values individually serves as a class or category.



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

Figure 2. Classifier classifying the unseen data[43]

III. METHODOLOGY

Search was made to find journal research articles which are published on the way from the period 2011-2018. The keywords were provided in such a way so as to find articles related to classification of structured clinical data but exclude time series data as the medical data related publications are too voluminous. The articles were searched from the database Pubmed as most of the reputed journals of Elsevier and Springer that are related to medical data mining are indexed in Pubmed. Search was done using two search queries as follows ('humans' was selected in the species filter):

Query 1: data mining classification clinical decision NOT survey NOT review NOT image NOT gene NOT genomics NOT protein NOT text NOT sensors NOT framework

Query 2: data mining classification clinical decision NOT survey NOT review NOT image NOT gene NOT genomics NOT protein NOT text NOT sensors NOT framework

Query 1 displayed 58 results whereas Query 2 displayed 96 results. Further screening was performed after reviewing abstracts and those articles which were not relevant to the objective of the study were rejected. Finally 41 articles were chosen to be reviewed.

IV. DATA_MINING AND MACHINE_LEARNING APPLICATIONS IN CLINICAL DECISION MAKING

The articles reviewed were classified into two categories: 1. Medicine and 2. Health care, based on the domain where the classification model can be adopted. 75% of the articles fell in the category medicine revealing the popularity of the domain among data mining researchers.

A. Medicine

Medicine incorporates the prognosis, diagnosis and the treatment given by doctor for curing of diseases. Researchers have employed various classification algorithms both "white-box" as well as "black-box" to solve problems related to medicine. White-box models were often used as they might be more preferable owing to their superior interpretability which is a highly desired quality for choosing models in medical field. Plentiful researches are done on the application of classification algorithms on disease data leading to the development of classifiers that can be utilised to diagnose diseases related to oncology, nephrology, endocrinology, neurology and so on. There were few publications on prediction of treatment as well. A brief explanation of a sample of works published in this sector is presented below.

Some researchers have published new rule based algorithm for diabetes classification [7]. A fitness function was introduced as a means of generating a comprehensive best rule set without affecting the diagnostic performance. The proposed approach has shown better, promising result compared to ID3, C4.5 and CART on the diabetes dataset in terms of diagnostic performance in 10-folded cross validation. The algorithm has also proved to outperform its counterparts as far as average rule length and average ruleset size are concerned.

A machine-learning module that derives classification rules to work alongside in a current knowledge base was presented [5]. The extracted rules detect the inappropriate prescription of the antimicrobial called piperacillin tazobactam. The proposed approach has shown reasonable results.

Clinical and environmental data had been analysed using algorithms on standard machine learning techniques in the study conducted [31]. In this exploratory and modelling study, library and field methods were employed to collect information. C5.0, CHAID, C &amp R models showed superior performance in case of clinical variables. In case of environmental variables, the suitable models identified for detection of lung cancer were C &amp and R models. Variables like pulmonary nodules, gush of plural fluid, size of pulmonary nodules, and site of pulmonary nodules have been identified as crucial variables that had the extreme effect on lung cancer detection. Data analysis was performed with Clementine 12 software.

Some researchers have recommended the use of support vector machine to predict axillary lymph node metastases in breast cancer patients [39]. The analysis was performed on the laboratory values of the main tumor. The dataset used consisted of the pathological data of 1325 breast cancer patients and the biological parameters included age, tumor size, grade and HER2. The proposed model performed with a diagnostic accuracy of 74.7%. Subgroup analysis with the parameters including luminal A, luminal B and HER2 subtypes showed inferior performance with 60% accuracy.

Some researchers have proposed fuzzy rules for treatment recommendation for patients suffering from breast cancer[29]. The patients were classified in either of the two classes: 1. Crisp treatment group and 2. Fuzzy treatment group. A dataset consisting of 471 patient data with HR(+) and HER2(-) breast cancer were analysed. The patients in the second group were additionally categorized into strata where patients in each stratum shared the identical predictive factor patterns showing analogous recurrence rates. 87.3% of HR(+) HER2(-) patients were allotted to second treatment group. Four diagnostic strata were created based on the survival tree model, and showed that patients with bad prognostic profiles inclined to receive endocrine therapy with chemotherapy. This suggested that postoperative chemotherapy is effective, though there was no significance statistically.

The application of an improved fuzzy rule based classifier to recognize diabetes was demonstrated [30]. The classifier referred to as c-means –neuro-fuzzy rule based classifier demonstrated better interpretability and accuracy on the UCI diabetes dataset. The complexity of the proposed approach was evaluated with respect to the number of fuzzy rules mined. Moreover the fuzzy rules extracted was similar to that of the rules used by medical experts in clinical practice. In comparison with the already published approaches in the literature, the fuzzy approach generates more interpretable, compact, and accurate classifier.

Çakır & Demirel have performed classification of breast cancer treatment methods based on the algorithms: IB1, multilayer perceptron and decision table [6]. Software called "Treatment Assistant" was developed to assist oncologists for the recommendation of treatment methods for breast cancer patients. 462 breast cancer patient data from Ankara Oncology Hospital were used to develop the treatment recommendation models for new patients. Learning algorithms were applied one by one on the dataset and the outcomes were compared to determine appropriate treatment method. The software uses the three algorithms to find the one that gives best result for each feature to detect the disease. IB1, Multilayer Perceptron and decision table show

best accuracy on hormonotherapy, tamoxifen and radiotherapy and chemotherapy outputs respectively.

B. Healthcare

Healthcare is the act of taking proactive or necessary care procedures to improve a patient's condition in and out of hospitals. The data mining applications in this domain include predicting the length-of-stay, predicting clinical complications, improving treatment outcomes and developing optimal practices all to maintain the wellbeing of the patient. The cost related to health care is too high [41]. There have been reasonable amount of publications in this sector. Some of the sample works published in this sector are explained below.

Researchers have conducted a study of adults admitted for sepsis in emergency department during October in the year 2013 to October in the year 2014 [34]. The dataset was split randomly into an 80:20 distribution for training and validation. A random forest model was built from data collected from electronic health records of four hospitals as a means of predicting in- hospital death. The proposed model was compared to logistic regression, classification and regression tree models and other models published in the literature and was proved to outperform its competitor models.

Zhu et al., have analyzed an HCUP state-wide inpatient discharge dataset that comprises patient demographics, care utilization and clinical data from California [40]. Records of heart failure Medicare beneficiaries with an inpatient experience of 11-months period were extracted. The data imbalance issue was rectified through under-sampling. In their study, logistic regression and decision tree algorithms were first applied to identify significant variables and to derive meaningful decision rules. Thereafter the original dataset was stratified accordingly and logistic regression was applied on each data stratum. The effect of dependent variables was further studied in the logistic regression modelling. Cross validation was performed and the results of conditional logistic regression were compared to that of standard classification models.

Faria et al., developed a system to estimate Quality of human Life using data mining. The clinical data of patients with cancer from ENT and Head and Neck services of an oncology institution, was analysed. The results shows that there were important variables (with respect to p-values) that contributed significantly to Quality of human life estimation of the patient. The crucial variables include size of the tumor and years of smoking. The optimal accuracy was achieved using the sequential minimal optimization algorithm developed by John and support vector machine was used for classification.

Ramezankhani et al., conducted a study on application of decision tree to identify low-risk population for Type-2 diabetes [27]. The Tehran_Lipid_and_Glucose_Study (TLGS) database consisting of 61 variables was used for the experiments. Important variables were chosen from demographic characteristics, drug and medical history and pathology measures. The accuracy of the proposed model was reported as 90.5%. The variables identified included triglycerides, educational level, body mass index, family_history of diabetes, fasting plasma glucose and job status.

Some researchers have proposed a machine-learning based algorithm to predict survival of patient after liver transplantation [11]. The model developed yielded promising results in accuracy, sensitivity, Kappa coefficient, and area under the curve. The proposed approach complements the current allocation system in efficiency and equity. The model minimises the prediction error produced by two other models considered for comparison.

Some researchers have proposed the use of a variation of support vector machine to predict complications related to percutaneous coronary intervention (PCI) [19]. The data BMC2 multicentre registry for the years 2007 and 2008 were used to construct models to predict thirteen various in-laboratory PCI complications. The performance of the developed models was evaluated and compared with the performance of three other classification algorithms. The proposed approach produced significant results over the other algorithms considered.

V. RESULTS AND DISCUSSION

Forty one articles were reviewed and the number of publications year-wise is depicted in Figure.3. Maximum amount of research have been published in 2015 and 2016. Least amount of works has been published in the year 2012. The quantity of research in 2017 has dropped. This may be due to the paradigm shift in research conducted from mining structured medical data to unstructured data like MRI, 3D Doppler, CT scan images etc.

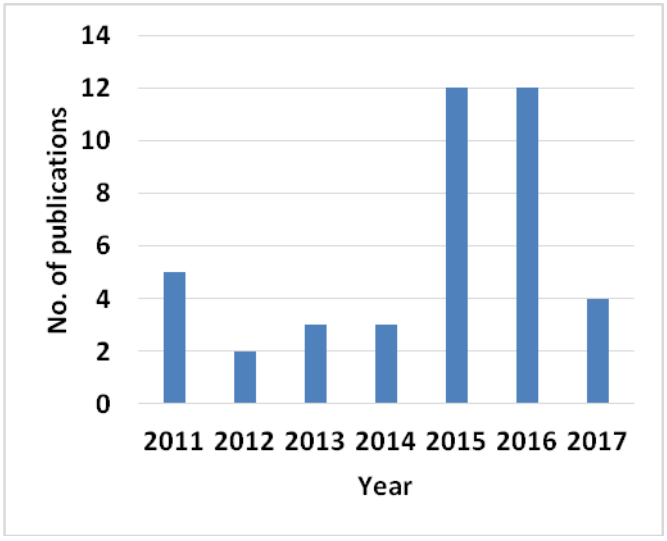


Figure 3. Year vs No. of publication

The articles chosen for this review belonged to various prominent publications some of them including Elsevier, Springer, Taylor & Francis and Wiley. The pie-chart (Figure. 4) depicts the distribution of articles among various publications. The outcome of the review indicates that approximately 37% of the articles were published by Elsevier, 10% of the articles were published by Springer and the remaining articles were published in IEEE, Taylor & Francis, Wiley, Biomed Central, IOS Press, Thieme Medical Publishers and others.

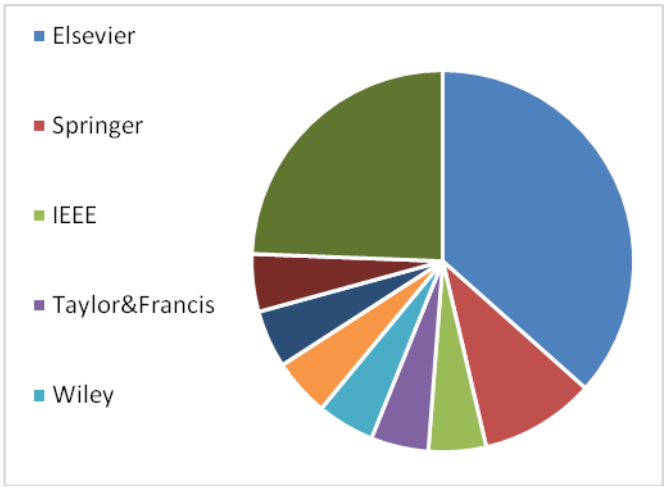


Figure 4. Distribution of articles published by various publishers

The articles chosen for this review were classified in one of the two categories: medicine and health care. Articles related to diagnosis and treatment were classified in the category 'medicine' and those that were related to overall care of in-patients and out-patients prior and after disease were classified in the category 'health care'. Approximately 75% of the articles fall under the category 'medicine' and the remaining articles fall under the category 'health care'. The distribution of articles in two categories are shown in Figure.5.

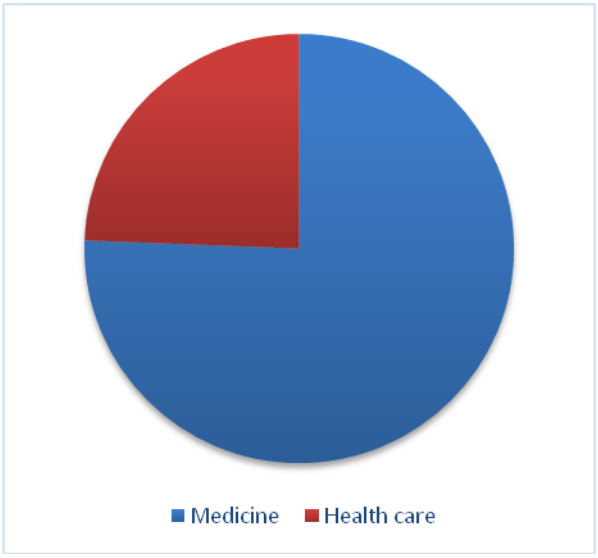


Figure 5. Distribution of articles in 'medicine' and 'health care'

The articles in medicine were related to prognosis, diagnosis and treatment of various diseases [16][17][18]. Predictive modelling has been implemented to diagnose diverse diseases including heart disease [28][40], kidney disease [18][33], diabetes [27], pulmonary disease [35], neurology related diseases [38] and various types of cancer [30][36][37]. Some researchers have applied machine learning techniques to provide diagnostic solutions that perform optimally in case of multiple diseases [9][14][21][26]

Most of the articles in this category were related to diagnosis and there were less research related to recommendation of treatment plan and thus would provide greater research scope for researchers in this area.

The articles in health care included the application of classification techniques in prediction of readmission of heart patients, prediction of survival of patients after kidney transplantation, prediction of need to refer post-operative patients to another hospital with better medical therapy and prediction of therapy complications [4][8][20][33][34].

Approximately 35% articles used black-box models like support vector machine, Naïve Bayes and neural networks to solve the problems [15][17]. Around 25% articles have employed white-box models like decision trees and rule-based classifiers [1][13][16] to provide solutions for various problems in both the categories. And in the rest of the articles both white-box and black-box algorithms have been employed [2][22][23][25]. Most of the articles reviewed have

adopted the standard classification algorithms such as support vector machine [39], decision trees [16] and logistic regression [40][12]. These algorithms have been very popular in the published literature. Decision tree tops them all because of its superior interpretability and efficient performance. Neural networks have also been used in used in good amount of studies but less when compared to the other three as it is relatively a new technique. In some of the studies, novel algorithms have been proposed [11][19][3][24]. Some researchers have developed hybrid models to solve certain problems in medical domain [10][32].

VI. RESEARCH GAP

Most of the research in medicine is related to prognosis and diagnosis of diseases and there has been less research in treatment recommendation and this study provides a direction for future research in this domain. Treatment recommendation will be a crucial topic as the decision support system can assist the doctors in deciding the optimal intervention, in choosing the optimal treatment plan and can also be beneficial for the medical students to have a better understanding of various treatments for diverse diseases. Choosing the best therapy and identifying the ideal treatment plan is a challenging task in case of diseases like heart failure and cancer. Moreover there is insufficient machine-learning research conducted in kidney diseases especially in chronic kidney disease and end-stage renal disease which are considered a global threat nowadays. Hence in spite of the growing trend in mining unstructured data, there is scope for further research in mining structured medical data and it must be given due importance as the pathology measures will enable inexpensive screening for diseases and can thereby contribute to early diagnosis of diseases and better health care.

VII. CONCLUSION AND FUTURE SCOPE

In this study, a systematic review on the application of classification techniques on medical data in the area of medicine and health care was conducted. There have been significant research published in these areas during the period 2011-2017, with maximum amount of works carried out in 2015 and 2016. 75% of the articles were related to 'medicine' and the remaining 25% of the articles were related to health care. Predictive modelling has been achieved mostly using standard algorithms and rarely using novel algorithms and the most popular standard algorithm being support vector machine. Most of the works published in medicine are related to prognosis and diagnosis and less research has been carried out in treatment recommendation which opens opportunities for further research.

REFERENCES

[1] Afzal, M., Hussain, M., Khan, W. A., Ali, T., Lee, S., Huh, E. N., & Hydari, M. A., *"Comprehensible knowledge model creation for cancer treatment decision making",* Computers in biology and medicine, Vol.82, pp.119-129.2017.

[2] Amaral, J. L., Lopes, A. J., Veiga, J., Faria, A. C., & Melo, P. L., *"High-accuracy detection of airway obstruction in asthma using machine learning algorithms and forced oscillation measurements"*, Computer methods and programs in biomedicine, Vol.144, pp.113-125, 2017.

[3] Armengol, E., *"Classification of melanomas in situ using knowledge discovery with explained case-based reasoning",* Artificial Intelligence in Medicine, 51(2), pp.93-105, 2011.

[4] Askland, K. D., Garnaat, S., Sibrava, N. J., Boisseau, C. L., Strong, D., Mancebo, M., & Eisen, J., *"Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy",* International journal of methods in psychiatric research, 24(2), pp.156-169, 2015.

[5] Beaudoin, M., Kabanza, F., Nault, V., & Valiquette, L., *"Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs",* Artificial intelligence in medicine, 68, pp.29-36, 2016.

[6] Çakır, A., & Demirel, B., *"A software tool for determination of breast cancer treatment methods using data mining approach",* Journal of medical systems, 35(6), pp.1503-1511, 2011.

[7] Cheruku, R., Edla, D. R., & Kuppili, V., *"SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification",* Computers in biology and medicine, 81, pp.79-92, 2017.

[8] Chia, C. C., Rubinfeld, I., Scirica, B. M., McMillan, S., Gurm, H. S., & Syed, Z., *" Looking beyond historical patient outcomes to improve clinical models"*, Science translational medicine, 4(131), 131ra49-131ra49, 2012.

[9] Christopher, J. J., Nehemiah, H. K., & Kannan, A., *"A swarm optimization approach for clinical knowledge mining",* Computer methods and programs in biomedicine, 121(3), pp.137-148, 2015.

[10] Chuang, C. L., *"Case-based reasoning support for liver disease diagnosis",* Artificial Intelligence in Medicine, 53(1), pp.15-23, 2011.

[11] Cruz-Ramirez, M., Hervas-Martinez, C., Fernandez, J. C., Briceno, J., & De La Mata, M., *"Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks",* Artificial intelligence in medicine, 58(1), pp.37-49, 2011.

[12] Dean, J. A., Wong, K. H., Welsh, L. C., Jones, A. B., Schick, U., Newbold, K. L., ... & Gulliford, S. L., *"Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy",* Radiotherapy and Oncology, 120(1), pp.21-27, 2016.

[13] Deng, Y., Groll, M. J., & Denecke, K., *"Rule-based Cervical Spine Defect Classification Using Medical Narratives",* Studies in health technology and informatics, 216, pp.1038-1038, 2015.

[14] Faria, B. M., Gonçalves, J., Reis, L. P., & Rocha, Á., *"A clinical support system based on quality of life estimation",* Journal of medical systems, 39(10), pp.114, 2015.

[15] Gaebel, J., Kolter, T., Arlt, F., & Denecke, K., *"Extraction Of Adverse Events From Clinical Documents To Support Decision Making Using Semantic Preprocessing",* Studies in health technology and informatics, 216, pp.1030-1036, 2015.

[16] Goodman, K. E., Lessler, J., Cosgrove, S. E., Harris, A. D., Lautenbach, E., Han, J. H., & Tamma, P. D., *"A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum β-Lactamase–producing organism",* Clinical Infectious Diseases, 63(7), pp.896-903, 2016.

[17] Guerrero, J. M., Martínez-Tomás, R., Rincón, M., & Peraita, H., *"Diagnosis of Cognitive Impairment Compatible with Early Diagnosis of Alzheimer's Disease",* Methods of Information in Medicine, 55(01), pp.42-49, 2016.

[18] Ichikawa, D., Saito, T., Ujita, W., & Oyama, H., *"How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach",* Journal of biomedical informatics, 64, pp.20-24, 2016.

[19] Lee, G., Gurm, H. S., & Syed, Z., *"Predicting complications of percutaneous coronary intervention using a novel support vector method",* Journal of the American Medical Informatics Association, 20(4), pp.778-786, 2013.

[20] Liu, X., Lu, R., Ma, J., Chen, L., & Qin, B., *"Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification",* IEEE journal of biomedical and health informatics, 20(2), pp.655-668, 2016.

[21] Nahato, K. B., Harichandran, K. N., & Arputharaj, K., *"Knowledge mining from clinical datasets using rough sets and backpropagation neural network",* Computational and mathematical methods in medicine, 2015.

[22] Ochab, M., & Wajs, W., *"Expert system supporting an early prediction of the bronchopulmonary dysplasia",* Computers in biology and medicine, 69, pp.236-244, 2016.

[23] Ozden, F. O., Ozgonenel, O., Ozden, B., & Aydogdu, A., *"Diagnosis of periodontal diseases using different classification algorithms: a preliminary study",* Nigerian journal of clinical practice, 18(3), pp.416-421, 2015.

[24] Parodi, S., Filiberti, R., Marroni, P., Libener, R., Ivaldi, G. P., Mussap, M., ... & Muselli, M., *"Differential diagnosis of pleural mesothelioma using Logic Learning Machine",* BMC bioinformatics, 16(9), S3, 2015.

[25] Prilutsky, D., Rogachev, B., Marks, R. S., Lobel, L., & Last, M., *"Classification of infectious diseases based on chemiluminescent signatures of phagocytes in whole blood",* Artificial intelligence in medicine, 52(3), pp.153-163, 2011.

[26] Rahulamathavan, Y., Veluru, S., Phan, R. C. W., Chambers, J. A., & Rajarajan, M., *"Privacy-preserving clinical decision support system using gaussian kernel-based classification",* IEEE journal of biomedical and health informatics, 18(1), pp.56-66, 2014.

[27] Ramezankhani, A., Pournik, O., Shahrabi, J., Khalili, D., Azizi, F., & Hadaegh, F., *"Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study",* Diabetes research and clinical practice, 105(3), pp.391-398, 2014.

[28] Ruiz-Fernandez, D., Torra, A. M., Soriano-Payá, A., Marin-Alonso, O., & Palencia, E. T., *"Aid decision algorithms to estimate the risk in congenital heart surgery",* Computer methods and programs in biomedicine, 126, pp.118-127, 2016.

[29] Santos, R. S., Malheiros, S. M., Cavalheiro, S., & De Oliveira, J. P., *"A data mining system for providing analytical information on brain tumors to public health decision makers",* Computer methods and programs in biomedicine, 109(3), pp.269-282, 2013.

[30] Settouti, N., Chikh, M. A., & Saidi, M., *"Generating fuzzy rules for constructing interpretable classifier of diabetes disease",* Australasian physical & engineering sciences in medicine, 35(3), pp.257-270, 2012.

[31] Shahhoseini, R., Ghazvini, A., Esmaeilpour, M., Pourtaghi, G., & Tofighi, S., *"Presentation of a model-based data mining to predict lung cancer",* Journal of research in health sciences, 15(3), pp.189-195, 2015.

[32] Somu, N., Raman, M. G., Kirthivasan, K., & Sriram, V. S., *"Hypergraph based feature selection technique for medical diagnosis",* Journal of medical systems, 40(11), pp.239, 2016.

[33] Tang, H., Poynton, M. R., Hurdle, J. F., Baird, B. C., Koford, J. K., & Goldfarb-Rumyantzev, A. S., *"Predicting three-year kidney graft survival in recipients with systemic lupus erythematosus",* ASAIO Journal, 57(4), pp.300-309, 2011.

[34] Taylor, R. A., Pare, J. R., Venkatesh, A. K., Mowafi, H., Melnick, E. R., Fleischman, W., & Hall, M. K., *"Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data–Driven, Machine Learning Approach",* Academic emergency medicine, 23(3), pp.269-278, 2016.

[35] Tuti, T., Agweyu, A., Mwaniki, P., Peek, N., & English, M., *"An exploration of mortality risk factors in non-severe pneumonia in children using clinical data from Kenya",* BMC medicine, 15(1), pp.201-210, 2017.

[36] Valdes, G., Solberg, T. D., Heskel, M., Ungar, L., & Simone II, C. B., *"Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy",* Physics in Medicine & Biology, 61(16), pp.6105-6112, 2016.

[37] Wang, Z., Wen, X., Lu, Y., Yao, Y., & Zhao, H., *"Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases",* Oncotarget, 7(11), pp.12612-12620, 2016.

[38] Weakley, A., Williams, J. A., Schmitter-Edgecombe, M., & Cook, D. J., *"Neuropsychological test selection for cognitive impairment classification: A machine learning approach",* Journal of clinical and experimental neuropsychology, 37(9), pp.899-916, 2015.

[39] Wu, J. L., Tseng, H. S., Yang, L. H., Wu, H. K., Kuo, S. J., Chen, S. T., & Chen, D. R., *"Prediction of axillary lymph node metastases in breast cancer patients based on pathologic information of the primary tumor",* Medical science monitor: international medical journal of experimental and clinical research, 20, pp.577-585, 2014.

[40] Zhu, K., Lou, Z., Zhou, J., Ballester, N., Kong, N., & Parikh, P., *"Predicting 30-day Hospital Readmission with Publicly Available Administrative Database",* Methods of information in medicine, 54(06), pp.560-567, 2015.

[41] O.A. Mohamed Jafar, "*A Study of Different Similarity Measures on the Performance of Fuzzy Clustering*", International Journal of Computer Sciences and Engineering, Vol.6, Issue.4, pp.168-173, 2018.

[42] N. Radha and S. Ramya , "*Performance Analysis of Machine Learning Algorithms for Predicting Chronic Kidney Disease*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.8, pp.72-76, 2015.

[43] Pier Luca Lanzi, "*Machine Learning and Data Mining*",Technology, Education, 2007.

Authors Profile

Mrs.Klinsega Jeberson is working as Assistant Professor in the department of Computer Science & I.T., Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, India. She has qualified National Eligibility Test (NET) in Computer Applications conducted by University Grants Commission (UGC India). Currently she is pursuing Ph.D. in Computer Applications at Sam Higginbottom universityof Agriculture, Technology and Sciences, Allahabad.

Dr. Manish Kumar is working as Associate Professor in the department of Information Technology, Indian Institute of Information Technology, Allahabad, India. His areas of interest are Wireless Sensor Networks, Data Management in WSN (Aggregation, Query Processing, Compressed Sensing, Routing), Distributed Database, Mobile Database System, Data Mining and Big Data Analytics.

Dr.Raghav Yadav is presently working as Associate Professor in the Department of Computer Science & Information Technology, at Sam Higginbottom University of Agriculture, Technology And Sciences. Prior to this, he has worked with Birla Institute of Technology (Extension Centre of BIT Mesra Ranchi), Allahabad and Indian Telephone Industries Limited (Government of India Undertaking) Naini, Allahabad.

