International Journal of Recent Development in Computer Technology & Software Applications
Vol. 5, Issue 1 – 2021
ISSN: 2581-6276

Application of Random Forest Algorithm in Healthcare Sectors

G. Haripriya[1], K. Abinaya[1], N. Aarthi[1], P. Praveen Kumar[1]
[1]*Department of Computer Science, IFET College of Engineering, Anna University, Villupuram, Tamil Nadu, India.*

Abstract

Safety culture is a multidimensional phrase in healthcare delivery systems that have been linked to medical errors and patient safety incidents. Despite this, there is little evidence to suggest that safety culture characteristics have an impact on overall patient safety. Aside from that, complex statistical analysis has only been applied sparingly in previous data studies on safety culture practices. In healthcare, there is a centralized organization delivery systems is a multifaceted issue that is linked to medical errors and patient safety concerns. Using health facility aggregate data from the United States, this study investigates the communication between recognized organizational safety components and patient security grade to address these safety concerns. Random forests computer vision program based on trees technique, researchers use to predict accurate and stable relationships between variables. Amidst this, there is little information available about which aspects of patient safety culture are most important. Furthermore, advanced statistical analysis in previous reviews of data on safety culture has been limited, as previously mentioned. Additionally, random forests, a machine learning technique based on trees, estimate precise and stable associations between variables. As a result, researchers used data from a study in collective US hospitals to investigate the relationship between identified safety culture components and patient safety grades. As a result of the findings, the scientists found that healthcare quality knowledge, organizational factors, and top management objectives all play a crucial influence in determining patient safety grades. Security concerns in the work unit, as well as the work environment created by hospital management, have an impact on patient safety outcomes.

Introduction

Healthcare facilities have been urged to establish safety management plans as an outcome of a greater comprehension of the significant cost and impact of medical errors. The sophistication of corporate safety culture has been linked to the success of these programs in researches[1]. As a result, it's critical to comprehend how safety culture fits into healthcare companies and drives them to grow and contribute to safety. In healthcare, safety culture refers to a collection of attitudes, beliefs, experiences, and expectations that describe a company's dedication to and effectiveness in implementing patient safety policies and procedures[2]. The most significant contribution of the research is the recommendation of a machine learning algorithm for use in PSC[3]. A significant contribution to this research is also made by the project's design, which makes extensive utilization of data from hospital-wide safety culture surveys to inform its development. The findings are intended to help hospitals in gaining valuable insight into the multifaceted safety development and its influence on patient safety by providing them with valuable information[4].

The remaining of the chapter is organized around this structure: Before defining the proposed random forest model, a review of the pertinent safety culture literature, evaluation computer vision and techniques methods is conducted. The model's results are displayed using data from hospital-level aggregates. The results of the model are shown using data from hospital-level aggregate surveys. Finally, we analyze the model and its findings, as well as its flaws, future research directions, and conclusions.

Concept of Patient Safety

The statement "how safety is managed over here" emphasizes the critical nature of considering how first impressions and ideas influence attitudes and behaviors regarding safety. Patient protection a few of the distinguishing features of high-quality healthcare, including timeliness, equity, efficacy, and a patient-centered approach[5]. The organization's security culture dictates its ability to establish continuous quality management objectives. A comprehensive patient safety evaluation aims to assist a business in gaining a scientific understanding of patient safety[6]. Identifying weak and robust safety cultures, tracking trends over time, reviewing organizational initiatives to promote patient safety, and comparing health departments and organizations are all important parts of this strategy. A strong safety culture has been associated to better patient outcomes in numerous healthcare settings[7]. Earlier studies in safety-sensitive industries attempted to link safety culture to operational safety, with variable degrees of success and mixed outcomes. Sorensen established a link between safety culture (or comparable attributes) and operational safety performance in order to better understand the relationship between the two[8]. He also developed acceptable performance metrics for determining improvements in safety culture and, as a result, operational safety performance. Since the early 2000s, the healthcare industry has developed and applied several strategies for analyzing and combining these two sections of patient safety.

Various surveys have been conducted utilizing the HSOPSC protection culture survey. ForReis and colleagues, for example, discovered that researchers and doctors all across the world were Patient safety culture is becoming more and more of a focus[9]. Between 2007 and 2016, 33 people died. The study included studies on HSOPSC from a variety of nations. Using the usual HSOPSC evaluation framework, studies rated the safety culture dimensions in their healthcare institutions as high or low. Once it relates to the protection of humans, the majority of studies that used HSOPSC indicated that the organization's culture is bad[10]. There are numerous systems, questionnaires, and evaluation procedures available for assessing an organization's safety culture, but quantifying the link between health and safety and safety record outcomes is not always straightforward[11][12]. It's also feasible that the multifaceted safety culture, with its potential interactions across features and importance to medication satisfaction in this study, is a factor will not be adequately defined given the current resources available in this research framework[13]. On the other hand, machine learning methods can provide a more precise assessment of PSC itself and impact patient safety because of their significant statistical analytic capabilities. In addition, such tools can aid in the ranking and interpretation of the value of the features utilized in the prediction[14]. As a result, utilizing machine learning approaches such as the random forest algorithm, the study of safety culture evaluations, the most prominent of HSOPSC, can benefit from creating more consistent, accurate, and versatile forecasts.

Random Forest Algorithms

Numerous industries and fields, including agriculture, transportation, electricity, and healthcare, have benefited from the attention-grabbing tree-based ensemble learning algorithms. These techniques have been used to forecast outcomes across a range of healthcare settings[15]. When designing a healthcare system, clinical outcomes, healthcare expenditures, and healthcare efficiency are considered [16]. Because these algorithms automatically handle interactions, they are also accurate predictors, even when a large number of variables are present. Random Forests (RFs) are a type of ensemble approach based on trees that is commonly used for classification and regression. Numerous recent studies have yielded reassuring empirical and theoretical findings.

Additionally, RFs are gaining traction in the healthcare industry, where they can be used for various purposes. A large number of decision trees are generated by RFs[17]. These trees are constructed by substituting variables from a randomly chosen subset of the original dataset for variables from the randomly chosen subset. RFs are capable of dealing with both categorical and numerical variables in prediction situations. One of the characteristics of RFs is their built-in cross-validation capability, which enables the independent variables to be ranked according to their association with the outcome variable, from most effective to least effective[18]. This simplifies the process of extracting functions from multisource data analysis. Breiman's presentation[19] provides an in-depth overview of radio frequency (RF) technology. While radiofrequency (RF) devices have the potential to automate a wide variety of operations on large and complex datasets, their utility in the context of particular safety culture has yet to be thoroughly investigated and evaluated. Researchers intend to build on the result of this research by combining an entirely new application field with a specialized

PSC background with a survey dataset to maximize radiofrequency (RF) technology benefits. Researchers intend to classify the most significant features (e.g., composites and qualities from the Survey) that influence patient safety grades in order to determine patient safety grades using the study approach described below.

Methodology of Study

Sources of Information

In 2016, the researchers compiled data from 677 annual staff surveys across the United States to create hospital-level data collection. The survey contains 42 variables and examines 12 composite safety culture ratings, comprising three to four separate survey items. The National Safety Council administers the survey[20]. The majority of questionnaires usea five-point Likert scale, with one signifying severe disagreement and five inferring strong consensus[21][22]. The expected score for each composite was determined using a percentage scale based on the average positive response rate (the proportion of agreeing and strongly agreeing responses to the total number of responses)[23]. On the other hand, respondents are asked to score their hospital's work area or unit on a scale of one to ten in the overall patient care rating, that researchers employed as a primary parameter. The outcome variable measured the average percentage of positive responses. Three additional hospital parameters were considered: categorized bed capacity, geographic location, and instruction status, both of which can affect patient care and safety. There are twelve safety culture composites and three hospital characteristics (categorical variables) listed in Table 1, among the independent variables. (variables that change throughout a period of time) The input was evaluated in two phases with the help of radiofrequency (RF) devices. The first stage investigated the significance of the properties of the 12 composites and hospital attributes in predicting patient safety grade in the hospital setting. Patient safety was investigated during Stage 2 by looking at the impact of 42 essential variables and institutional characteristics.

Table 1.Grid search evaluation factor search space

Grid Parameters	Ranges
Maxi_intensity	2,3,4,6,7,8,9,10,15
Minimum_pattern__per_clover	2,3,4,5,15
Minimum_pattern_per_partition	2,3,5,10,15
N_Grid_predictor	200,400,600,700,800,1000,1500

Random forest Algorithm

Every version divides statistics into two groups at random: a schooling set (which comprises 80 percent of the sample) and a trying-out set (which comprises the remaining 20 percent of the sample) (the closing 20 percent). (20 percent of the total sample size). At this point, the designers went over the hyper-parameters that are commonly used in RF algorithms, which included the following: There are n estimators (different types of wood) inside the forest:

- max intensity (the highest level of intensity in the tree)
- min pattern partition (the smallest quantity of statistical factors in a tree prior the node splits)
- min samples leaf (the smallest number of statistics factors in a tree before the node is split)

By subjecting the algorithm to an exhaustive grid scan, the hyper-parameters are evaluated to improve the algorithm's accuracy. Grid search identifies the most influential parameter combinations for tuning purposes during the gadget learning process. Table 2 summarizes the parameter seek an area utilized within the grid seek to take a look at the environment. Several metrics were used to analyze and interpret the predictive output of each version after the grid seek analysis, including the mean absolute percent error (MAPE), the mean absolute error (MAE), and the mean rectangular error (MRE).

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - p_i| \,/\, y_i \qquad (1)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - p_i| \qquad (2)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - p_i|^2 \qquad (3)$$

Where, y_i is the real value in the i^{th} observation and p_i is the estimate in the similar observation within the dataset size.

Models 1 and 2 were fitted in parallel with the previously stated Stages 1 and 2. i.e., There are various techniques and styles that have been chosen for interlaced and complex variables (Models 1) in parallel with the previously mentioned Stages 1 and 2. (Models 2). Figure 2 shows a model of the system. When tuning hyper-parameter combinations using grid search for composites, Model 1 shows the effects on composites, whereas Model 2 shows the same for individual variables. A value is assigned to each feature to graphically represent the relative relevance of each statistic, which is an essential function of RFs. For this, the Scikit-Learn[24] module in RFs calculates the value of a function by examining the level of impurity to be reduced on average by tree nodes that employ it. The weight of each and every node is determined by the number of training samples used. The score for each function is calculated automatically after training is completed. As a result, the total number of characteristics is used to scale the findings.

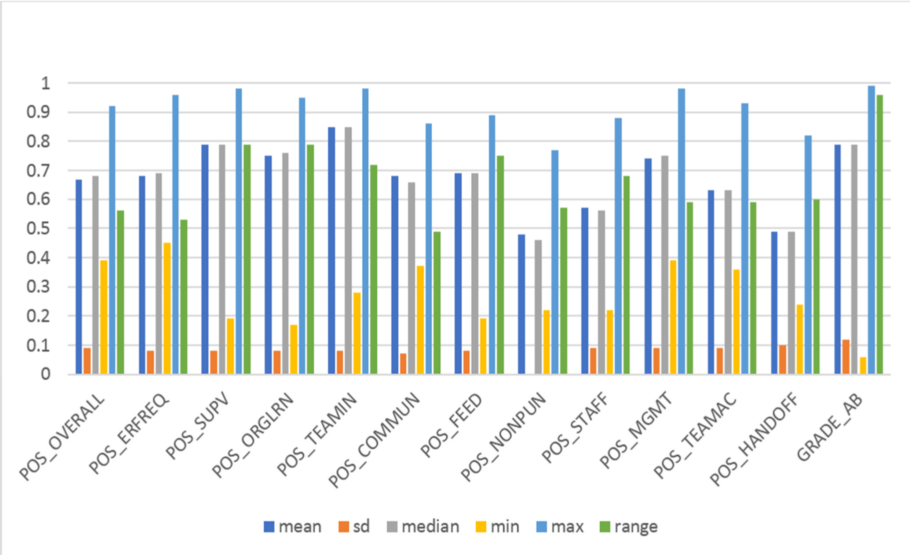

Figure 2.Descriptive summary of study composites and covariates

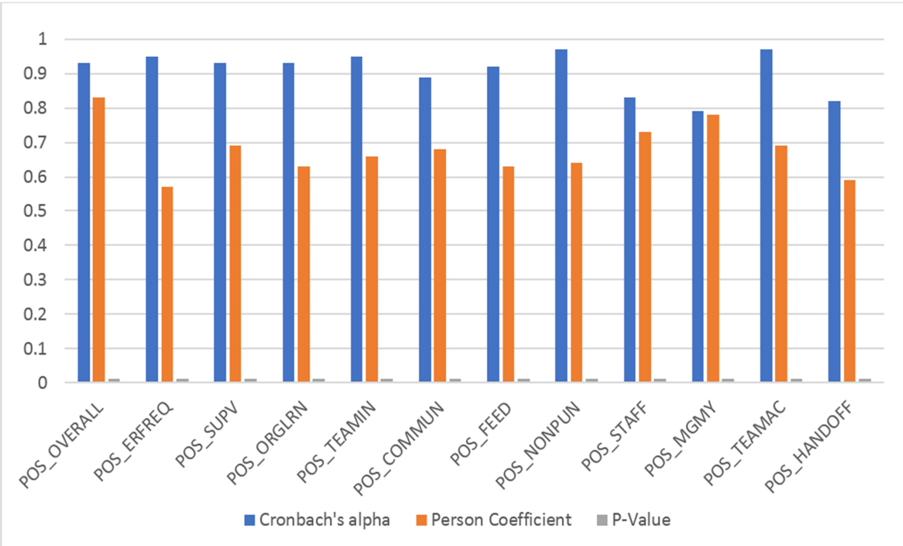

Figure 3.The frequency of the correlation among continuous parameters and the sufferer privacy level

Performance Analysis

The goal of this chapter is to figure out which composite scores and individual safety culture components [25][26] are the most closely linked. Overall patient safety has been assigned a grade. Prior to the study, descriptive statistics were employed to classify missing data and possible distributional outliers. As a result, they were left out of the final report, leaving only 672 people who took part. For the measured composites, the descriptive statistics[27][28][29] are easily accessible here: The figure's mean value is used to determine the average percent. The statistical features of the data were determined by a series of linear correlation experiments with composites (continuous variables) and the patient safety grade. Patient safety is typically regarded as good: processes and systems efficiently prevent errors, but patient safety concerns are regarded as inadequate. Help with administration: The hospital management promotes a safe working environment for patients and priorities personnel safety.

- Encouraging employees to identify ways to increase patient security.
- Praising employees for following up on patient care practices are just a few of the goals for supervisors and managers. Not overlooking patient safety concerns, among other things.

Conclusion

Health agencies need to learn how and create a safety culture to reap the benefits of enhanced safety. In this investigation, a high level of relative importance was found for the composite culture based on patient averages, and individual defensive sample factor grades using a random HSOPSC forest method. As known, this is the first time such evaluation has been conducted. In terms of safety cultivation, random forest algorithms are regarded safe. The first data analysis suggests that management assistance (37%) and supervisor and management support (63%) are essential composites that are of low and high value in most hospitals. This chapter shows that the safety culture can be improved by showing the relative importance of specific aspects of the culture of patient care measured by the average level of employees. As grid search analysis has been conducted in the same RF model, estimation accuracy may be increased to refine the hyper parameter. It is anticipated that future research will concentrate on parameter tuning optimization and will be capable of evaluating different types of motion. Organizational culture and health and care afferent are reflected in organizations' safety culture, which was accomplished through the inclusion of the necessary details for each specific industry. It is unclear how widely our findings in the field of healthcare are replicated in other countries, particularly in the developing world. Health organizations make significant investments in the field and possess the resources necessary to expedite patient protection. For health institutions to reap the benefits of these investments, they need to be aware of the critical aspects of a security culture that have been identified in this investigation. The use of the technology can be advantageous for future study in a certain context. Staff in a hypothetical hospital will profit from individual survey findings as the basis of information.

References

1. G. Sasubilli and A. Kumar, "Machine Learning and Big Data Implementation on Health Care data," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 859-864, doi: 10.1109/ICICCS48265.2020.9120906.
2. S. A. Kumar, H. Kumar, V. Dutt and H. Soni, "Self-Health Analysis with Two Step Histogram based Procedure using Machine Learning," 2021 Third International

Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 794-799, doi: 10.1109/ICICV50876.2021.9388427.

3. S. R. Swarna, S. Boyapati, V. Dutt and K. Bajaj, "Deep Learning in Dynamic Modeling of Medical Imaging: A Review Study," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 745-749, doi: 10.1109/ ICISS49785.2020.9315990.
4. S. A. Kumar, A. Kumar, V. Dutt and R. Agrawal, "Multi Model Implementation on General Medicine Prediction with Quantum Neural Networks," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 1391-1395, doi: 10.1109/ICICV50876.2021.9388 575.
5. Al-Ahmadi, Talal A. "Measuring Patient Safety Culture in Riyadh's Hospitals: A Comparison between Public and Private Hospitals." The Journal of the Egyptian Public Health Association 84, no. 5-6 (2009): 479-500.
6. S. A. Kumar, H. Kumar, V. Dutt and H. Soni, "Self-Health Analysis with Two Step Histogram based Procedure using Machine Learning," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 794-799, doi: 10.1109/ICICV50876.2021.9388427.
7. Al-Mandhari, Ahmed, Ibrahim Al-Zakwani, Moosa Al-Kindi, JihaneTawilah, Atsu SS Dorvlo, and Samir Al-Adawi. "Patient safety culture assessment in Oman." Oman medical journal 29, no. 4 (2014): 264.
8. Swarn Avinash Kumar, Harsh Kumar, SrinivasaRaoSwarna, Vishal Dutt, "Early Diagnosis and Prediction of Recurrent Cancer Occurrence in a Patient Using Machine Learning", European Journal of Molecular & Clinical Medicine, 2020, Volume 7, Issue 7, Pages 6785-6794.
9. Kumar, K. Suresh, T. Ananth Kumar, S. Sundaresan, and V. Kishore Kumar. "Green IoT for 9 Sustainable Growth and Energy Management in Smart Cities." Handbook of Green Engineering Technologies for Sustainable Smart Cities (2021): 155.
10. S. Boyapati, S. R. Swarna, V. Dutt and N. Vyas, "Big Data Approach for Medical Data Classification: A Review Study," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 762-766, doi: 10.1109/ ICISS49785.2020.9315870.
11. Vishal Dutt, RohitRaturi, Vicente García-Díaz, SreenivasSasubilli, "Two-Way Bernoulli distribution for Predicting Dementia with Machine Learning and Deep Learning Methodologies", Solid State Technology, 63(6), pp.: 9528-9546.
12. SwarnAvinash Kumar, Harsh Kumar, Vishal Dutt, HimanshuSwarnkar, "Role of Machine Learning in Pattern Evaluation of COVID-19 Pandemic: A Study for Attribute Explorations and Correlations Discovery among Variables", (2020): Global Journal on Application of Data Science and Internet of Things, Vol 4 No 2, [ISSN: 2581-4370].
13. Biau, Gérard, Luc Devroye, and Gäbor Lugosi. "Consistency of random forests and other averaging classifiers." Journal of Machine Learning Research 9, no. 9 (2008).

14. Vishal Dutt, Sriramakrishnan Chandrasekaran, Vicente García-Díaz, (2020). “Quantum neural networks for disease treatment identification.”, European Journal of Molecular & Clinical Medicine, 7(11), 57-67.
15. Vikas Kumar Singh, Dr. Sanjay Pawar, LohitShekam, Vishal Dutt (2020),” Impact Of Covid 19 On Fmcg Sector.” Journal of Critical Reviews, 7 (12), 4477-4484. doi:10.31838/jcr.07.12.640.
16. SwarnAvinash Kumar, Harsh Kumar, Vishal Dutt, Himanshu Swarnkar, “Contribution Of Machine Learning Techniques To Detect Disease In-Patients: A Comprehensive Analysisof Classification Techniques”, Global Journal on Innovation, Opportunities and Challenges in AAIand Machine Learning, Vol. 3, Issue 1 -2019, ISSN: 2581-5156.
17. Cooper, M. D., M. Collins, R. Bernard, S. Schwann, and R. J. Knox. "Criterion-related validity of the cultural web when assessing safety culture." Safety science 111 (2019): 49-66.
18. Swarn Avinash Kumar, Harsh Kumar, Vishal Dutt, Pooja Dixit, “The Role of Machine Learning in COVID-19 in Medical Domain: A Survey”, Journal on Recent Innovation in Cloud Computing, Virtualization & Web Applications, Vol 4 No 1 (2020), [ISSN: 2581-544X].
19. Fujita, Shigeru, KanakoSeto, Takefumi Kitazawa, Kunichika Matsumoto, and Tomonori Hasegawa. "Characteristics of unit-level patient safety culture in hospitals in Japan: a cross-sectional study." BMC health services research 14, no. 1 (2014): 1-9.
20. SwarnAvinash Kumar, Harsh Kumar, Vishal Dutt, Pooja Dixit, “Deep Analysis of COVID-19 Pandemic using Machine Learning Techniques”, (2020): Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning, Vol 4 No 2, [ISSN: 2581-5156].
21. Gurm, Hitinder S., Judith Kooiman, Thomas LaLonde, Cindy Grines, David Share, and Milan Seth. "A random forest based risk model for reliable and accurate prediction of receipt of transfusion in patients undergoing percutaneous coronary intervention." PloS one 9, no. 5 (2014): e96385.
22. SwarnAvinash Kumar, Harsh Kumar, Vishal Dutt, Himanshu Swarnkar, “COVID-19 Pandemic analysis using SVM Classifier: Machine Learning in Health Domain”, Global Journal on Application of Data Science and Internet of Things, 2020, Vol 4 No. 1.
23. S.Sasubilli ,A. Kumar,V.Dutt, "Machine Learning Implementation on Medical Domain to Identify Disease Insights using TMS", 2020, Sixth International Conference on Advances in Computing & Communication Engineering Las Vegas USA ICACCE 2020 (22-24 June) ISBN: 978-1-7281-6362-8.
24. Itoh, Kenji, Henning Boje Andersen, and Kim Lyngby Mikkelsen. "Safety culture dimensions, patient safety outcomes and their correlations." In Patient Safety Culture, pp. 67-98. CRC Press, 2018.
25. Abhishek Kumar, TvmSairam, Vishal Dutt, “Machine Learning Implementation For Smart Health Records: A Digital Carry Card”, Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning Vol. 3, Issue 1-2019.

26. McWilliams, Christopher J., Daniel J. Lawson, Raul Santos-Rodriguez, Iain D. Gilchrist, Alan Champneys, Timothy H. Gould, Mathew JC Thomas, and Christopher P. Bourdeaux. "Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK." BMJ open 9, no. 3Najjar, Shahenaz, ElfiBaillien, Kris Vanhaecht, MotasemHamdan, Martin Euwema, Arthur Vleugels, Walter Sermeus, Ward Schrooten, Johan Hellings, and AnnemieVlayen. "Similarities and differences in the associations between patient safety culture dimensions and self-reported outcomes in two different cultural settings: a national cross-sectional study in Palestinian and Belgian hospitals." BMJ open 8, no. 7 (2018): e021504. (2019): e025925.
27. S. M. Sasubilli, A. Kumar and V. Dutt, "Improving Health Care by Help of Internet of Things and Bigdata Analytics and Cloud Computing," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, 2020, pp. 1-4, doi: 10.1109/ICACCE49060.2020.9155042.
28. R. Raturi and A.Kumar " An Analytical Approach for Health Data Analysis and finding the Correlations of attributes using Decision Tree and W-Logistic Modal Process", 2019, IJIRCCE Vol 7, Issue 6, ISSN(Online): 2320-9801 ISSN (Print) : 23209798.
29. S. Chandrasekaran and A.Kumar Implementing Medical Data Processing with Ann with Hybrid Approach of Implementation Journal of Advanced Research in Dynamical and Control Systems-JARDCS issue 10, vol.10, page 45-52, ISSN-1943-023X. 2018/09/15.

