Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7654194

HEALTHCARE FRAUD DETECTION USING MACHINE LEARNING AND SENTIMENT DATA ANALYSIS

M. Jayanthi Rao[1], D. Apparao[2], B. Ramesh Naidu[3], B. Ramakrishna[2*], M. Ramanaiah[4] and M. Balakrishna[5]

1. Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali-532201, India.
2. Department of Mechanical Engineering, Aditya Institute of Technology and Management, Tekkali-532201, India.
3. Department of Information Technology, Aditya Institute of Technology and Management, Tekkali-532201, India.
4. Department of Chemistry, Aditya Institute of Technology and Management, Tekkali-532201, India.
5. Department of Chemistry, Lendi Institute of Engineering and Technology, Vizianagaram-535005, India.

* Corresponding author: brkbtech@gmail.com

Abstract

Clinical judgment is important in social classes, and it should be reasonable. There are many moving pieces in the complex system that is the clinical consideration business. Rapid expansion is taking place. At the same time, deceit in this industry is becoming a major problem. The mistreatment of the clinical insurance structures is one of the problems. In the clinical benefits sector, manually identifying cheaters is a taxing task. Recently, AI and data analytics techniques have been utilized to reliably identify clinical benefits frauds. In this paper, we want to provide a general overview of clinical benefits sector fraud as well as detection techniques With gratitude Diverse open investigations were taken into consideration in the writing task with a complement on the procedures used, choosing the enormous sources, and the characteristics of the clinical benefits data. The general AI strategies and from almost immediately obtained sources of clinical consideration data would be approaching subjects crucial to make clinical consideration sensible, to weaken the reasonability of clinical benefits coercion disclosure, and to introduce topnotch clinical consideration systems, it can be inferred from this overview. As discussed in this study, numerous new investigations apply AI and data analytics to identify distortion in the clinical care business. Additional research is required to choose certain bizarre occurrences. More recent AI techniques and hospital quality evaluation using machine learning and data analysis of patient abuse can be employed to further promote outcomes

Keywords: Data analysis, achine learning, Artificial intelligence, and Clinical Consideration

Introduction

Sentiment analysis is a technique for detecting and extracting significant subjective information from text, visuals that reflect emotion or sentiment and multi-modal data[1]. It is often a positive, neutral, or negative assessment of an object of interest[2]. It can also be used to discover the

user's opinions, thoughts, emotions, appraisals, feelings, attitudes, traits, and behaviour towards products, services, organisations, individuals, issues, events, and themes. During the global covid-19 outbreak, internet usage drastically rose (India by more than 40 percent). People are utilising and participating with electronic media to express criticism or review, engage in discourse on forums, blogs, e-commerce rating, Twitter and LinkedIn, WhatsApp, and other social networking media, and to provide e-commerce ratings and ratings. Sentiment analysis is a developing subject of natural language processing research (NLP). Opinions have a crucial influence in human endeavours. In general, their likes and dislikes dictate their actions and routines. For this reason, whenever deciding, they typically seek the opinions, experiences, and perspectives of others. This is a frequent occurrence among both individuals and businesses[3]. Classifications based on granularity include document, phrase, word, and character levels. Classification according to method as supervised, unsupervised, and semi-supervised learning [4] A classification of emotions based on textual features. This may comprise words, their occurrence, grammatical representation and classification, the presence of sentiment words, phrases in sentences, language norms, the sentiment changer, and syntactic dependency[5].

Clinical benefits have and support to be a fundamental part in social classes lives. The human body is a compound development. In this manner, it is major to have master specialists ready to break down and treat afflictions in different bits of the body. This actuates a couple of sorts of treatment procedure that specialists complete for patients in different specialties. The place of the prosperity business is to viably fill in whatever number patients as could be anticipated considering the present situation. However, with every treatment there is an expense related with each help gave. Specialists, road drug specialists and clinical staff should be paid for their time and capacity including diverse clinical comforts. Every now and again these expenses are not sensible to the patients. Thusly, insurance plans are used to allocate costs over all patients in the clinical consideration structure and pay for the basic people and stuff. In like manner with any assurance system, there is a chance for misuse or blackmail works out. Clinical consideration coercion is dynamically apperceived as one of certifiable social concerns. Clearly, clinical consideration coercion is an issue for the public power and there is a necessity for convincing ID strategies.

To perceive clinical consideration blackmail, it requires part of attempts with wide clinical data. Usually, clinical consideration coercion disclosure tremendously depends upon the experience of region subject matter experts, which is enough mixed up, expensive and monotonous. Manual recognizable proof of clinical consideration deception incorporates several analysts who review and recognize the questionable clinical security claims which requires a great deal of effort. Regardless, the state-of-the-art advances of AI and data analytics techniques incited more useful and automated area of clinical benefits swindles. There has been a creating interest in burrowing clinical benefits data for distortion ID in the new years. This paper reviews the various strategies used for recognizing the underhanded activities in Health security ensure data. Below architecture (Figure-1) represents three dimensions of hospital performance.

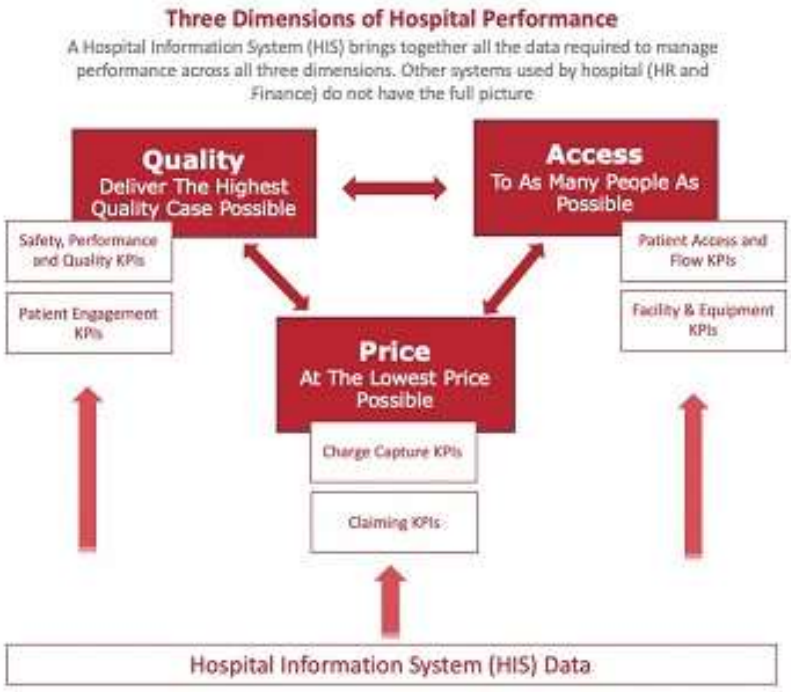


Figure.1 Hospital Quality Evaluation

Literature Survey

Motivation

According to M.N. Sarkies et al[6], manual data collection from ward-based sources obtained only 376 (69%) of the 542 inpatient episodes acquired by the hospital administrative electronic patient management technology. The highest levels of agreement were reported between administrative data from the computerized patient management program and inpatient medical record review for length of stay (93,4%) and discharge destination (91%) data. Qi Liu and Miklos Vasarhelyi[7] previously reported some preliminary knowledge of the US health care system and its fraudulent behaviors, analyzed the characteristics of health care data, reviewed and compared fraud detection approaches using health care data from the literature as well as their corresponding data preprocessing methods, and proposed a geo-location clustering model. R.A. Bauder and T.M. Khoshgoftaar[8] suggested a broad Bayesian inference-based probabilistic programming model for outlier detection. Unlike most outlier detection methods, our model delivers probability distributions as opposed to simple point values. Credible intervals are also created to increase the certainty that the detected outliers are, in fact, outliers. S. Altuntas, T. Dereli, and Z. Erdoan[9] proposed a service quality evaluation model based on the service quality measurement scale and machine learning algorithm. The SERVQUAL scale is primarily used to determine items impacting service quality. Following that, a service quality assessment model is created to efficiently manage the available resources in order to improve operations.

Existing System and Its Limitations

Clinical consideration distortion acknowledgment inconceivably depends upon the experience of region trained professionals, which is adequately mixed up, exorbitant and drawn-out. Manual acknowledgment of clinical consideration distortion incorporates a few inspectors who

truly review and perceive the questionable clinical security claims which requires a ton of effort. In any case, the state-of-the-art advances of AI and data mining techniques provoked more powerful and mechanized revelation of clinical benefits swindles. Manual detection of healthcare fraud needs a few auditors to manually evaluate and identify questionable medical insurance claims, which takes a lot of time and work.

Proposed System and Its Advantages

Statistical indicators derived from machine learning algorithms are among the most prominent tools for gauging the success of hospitals, alongside regulatory inspections, public opinion polls, independent evaluations, and other techniques of evaluation. We propose a general peculiarity acknowledgment model, taking into account Bayesian derivation, using probabilistic programming. Our model gives probability transports rather than just pointing values, similarly with most typical exemption distinguishing proof techniques. Reasonable ranges are in like manner created to extra overhaul sureness that the perceived special cases should without a doubt be seen as inconsistencies. Two logical examinations are presented displaying our model's sufficiency in perceiving abnormalities. The essential logical investigation uses temperature data to give a sensible assessment of a couple of abnormality ID strategies. In order to detect instances of healthcare fraud, a significant amount of work and an in-depth understanding of relevant medical topics are required, together with the application of ML, data analytics and artificial intelligence techniques[10-14]. The proposed model represents in Figure-2.

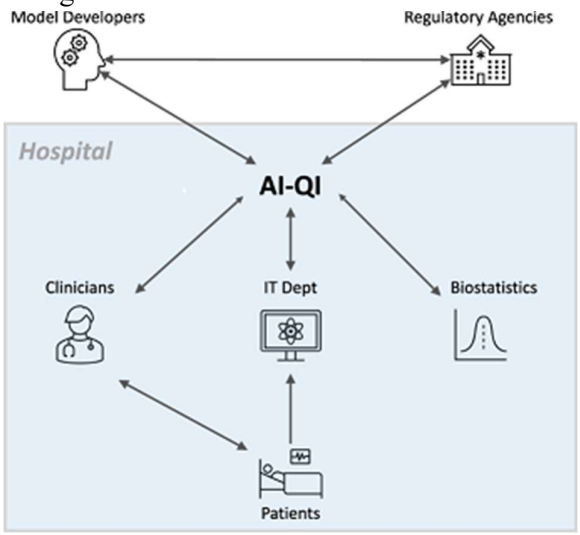


Figure 2. Machine Learning Development Process

Methodology

Sentiment Analysis[15]

Machine learning is not limited to the realms of business and commerce; it also finds widespread use in fields like healthcare and medicine. Patient care, stroke prediction, cardiology, and even individual health studies have all benefited from the use of machine learning. In addition, sentiment analysis and text categorization are two common ways that machine learning is utilised to assess patient experience feedback. For text and sentiment

analysis, the most popular methods were unsupervised learning and supervised learning[16, 17]. Supervised learning, in which a subset of data is manually categorises based on themes and sentiment, was the most popular approach. In Table-1 below, you will find a compilation of reviews culled from various patient perspectives shared via social media.

Table-1: Sentiment Data Analysis

Categorization	Summarization	Analysis Sample
Positive	Expressions of liking, approval, and gratitude	In general, I'm pleased with my experience at this facility. The staff at the hospital is friendly and accommodating.
	The good points of hospital care and infrastructure	The time we had to wait was quite short. The staff at the pharmacy counter did a fantastic job.
	Positive qualities of staff	The staff is courteous and helpful.
	Encourage or advise that others adopt	I strongly suggest that you give birth at this particular medical facility.
	Service's beneficial/desirable effects	I'd want to express my gratitude to you, Mr. X, for performing colon surgery on my dad. He is currently enjoying excellent health.
Negative	Dislike or disapproving expression	That guard makes me so angry. What a jerk he was to me! Moreover, the food is tasteless.
	A negative aspect of hospital services or facilities	The process for discharging patients was painfully slow. Obtaining a parking spot is challenging because there are so few of them
	employees with negative traits	Staff nurses were impolite and obstinate. I asked for help but received no response. The doctor scolded us for seeking care at 3 a.m. in the emergency room. We were irritated by his behavior.

	Unfavorable and unintended consequences	My parents left my dad alone in the bathroom for a few minutes after he fell in the toilet. Our family requires an explanation from the hospital director.
Neutral	Review that provides information with no opinion.	The particular is one of the good cardiac center.
	Examine as questions	Is there a spine surgeon at your hospital?
	Too confusing/not clear/ Only greetings to offer	Have a nice time. No comment. First, let's watch and see what happens.

Evaluation of Machine Learning Performance

SVM and NB are the most used classifiers in supervised learning. SVM and NB are the most often employed classifiers in supervised learning, and both consistently provide excellent classification performance. Topic modelling, on the other hand, is an unsupervised machine learning method that uses Latent Dirichlet Allocation (LDA) to automatically identify subjects within a given comment. LDA is a text generation model based on the premise that a document's words represent a collection of latent themes.

Figure 3 depicts the metrics that can be used to evaluate the effectiveness of machine learning algorithms: accuracy, sensitivity, recall, specificity, precision, hamming loss, and the F-measure. The F1 score of the model indicates its quality. This study's research revealed sentiment analysis using the NB classifier. Using the NB classifier, positive sentiment analysis has a score of 0.97, negative sentiment analysis has a score of 0.59, and the average text classification score is 0.78%. In contrast, when the NB classifier was applied to a study of patient satisfaction at one health Care System, the emotion score was 0.80 and the text score was 0.73. According to a separate study, the patient feedback had an emotion score of 0.88, a theme score of 0.90 for dignity and respect, and a text classification score of 0.88. To ensure purity, utilise the NB algorithm. Despite this, an SVM-based machine learning analysis of review comments generated a score of F1.

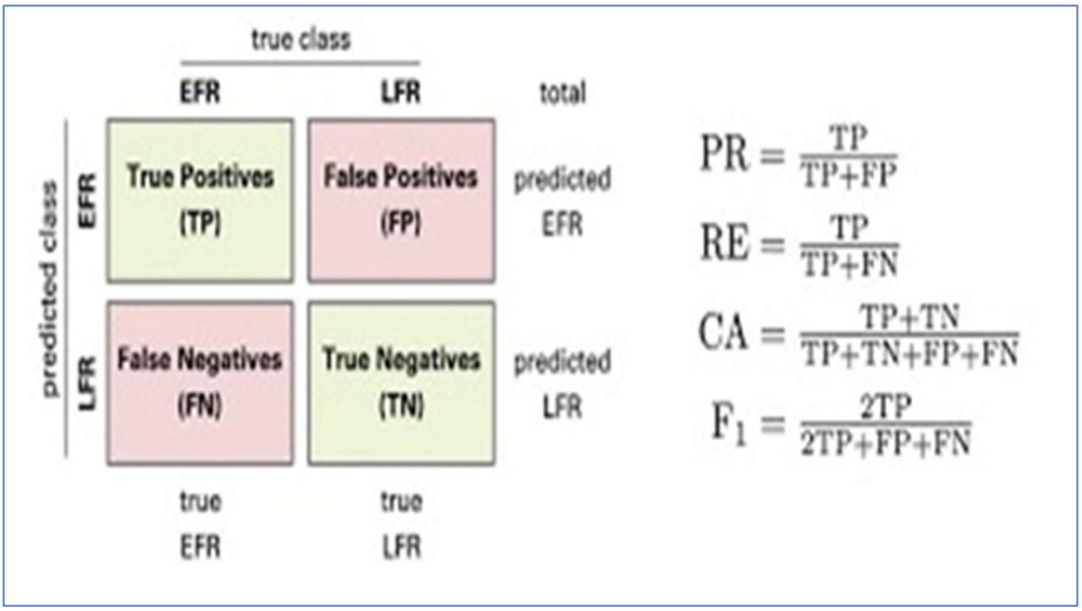


Figure.3 Performance Evaluations Metrics

Experimental Analysis and Results

Tensorflow

Tensor Flow is a free and open-source framework for data flow and differentiable programming across a variety of activities. It is an important numerical library that is also utilized for AI applications such as neural connections. Google use it for both research and development. Tensor Flow was developed by the Google Brain team for internal usage. It was transmitted according to the Apache 2.0open -source grant on November 9, 2015.

Numpy

Numpy is an extensively valuable display taking care of pack. It gives a predominant presentation multidimensional group thing, and mechanical assemblies for working withthese displays. It is the fundamental pack for sensible figuring with Python. It contains various features including these huge ones.

1. An amazing N-dimensional show object
2. Refined (telecom) limits
3. Gadgets for organizing C/C++ and Fortran code
4. Significant direct factor based math, Fourier change, and subjective number limits

Other than its obvious sensible uses, Numpy can similarly be used as a capable multi-dimensional compartment of customary data. Self-decisive data types can be described using Numpy which licenses Numpy to consolidate with a wide collection of informational indexes reliably and conveniently.

Pandas

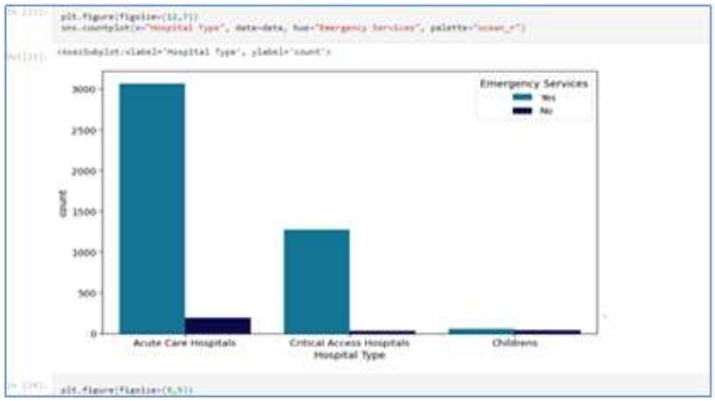
Pandas is an open-source Python library that provides first-rate execution data management and analysis tools via its wonderful data structures. Python was mostly utilized for data manipulation and organization. It had virtually no responsibility for data analysis. Pandas solved this problem. Using Pandas, we may do five standard processes in the management and

evaluation of data, ignoring the order of data load, prepare, control, model, and analyze. Python using Pandas is used in a variety of academic and business domains, including finance, monetary perspectives, statistics, evaluation, etc.

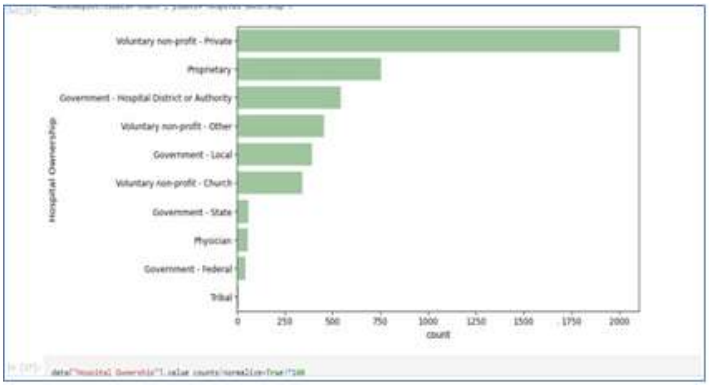
Matplotlib

Matplotlib is a Python 2Dplotting package that generates publication-quality figures in a set of printed duplicate plans and intelligent conditions across stages. Matplotlib is compatible with Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four GUI tools. Matplotlib strives to make difficult tasks simple and straightforward. Several lines of code can generate graphs, histograms, power spectra, bar charts, cluster diagrams, scatter plots, etc. For models, please refer to the model plots and thumbnail display; for direct charting, the pyplot module provides a MATLAB-like interface, especially when combined with IPython. For the power user, you have complete control over line styles, text style characteristics, hatchet properties, etc. via an article-based interface or through a number of MATLAB-specific limitations.

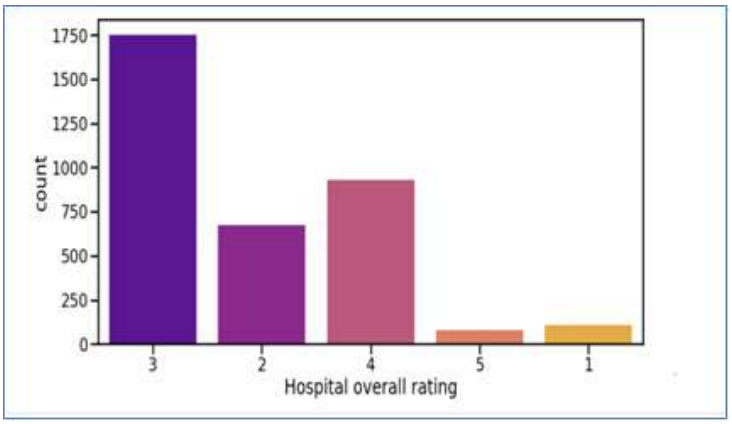
Graph Analysis



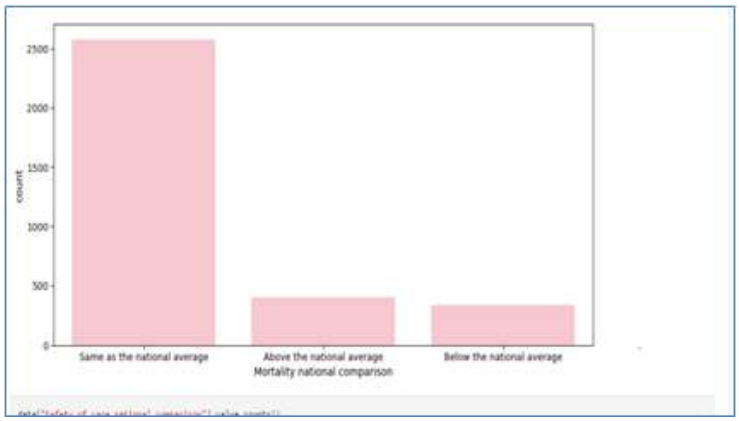
Emergency services graph



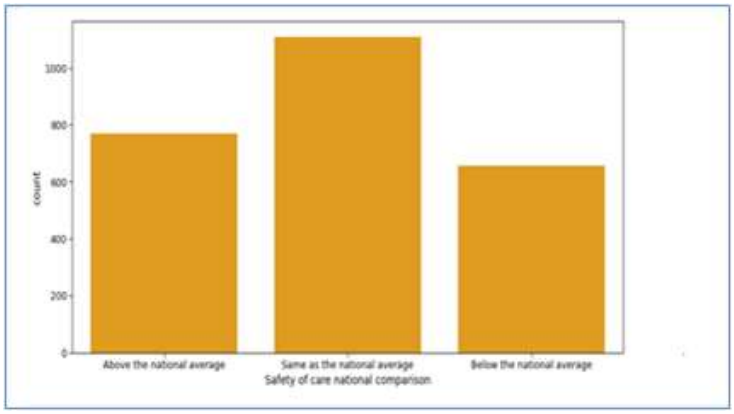
Hospital ownership count graph



Hospital overall rating graph



Hospital average counting graph



Hospital safety comparison graph

Conclusion

In this paper, techniques for clinical benefits fakes, types and sources of clinical benefits data, and clinical consideration blackmail were considered. The composition surveys a number of examinations. Data is argued to be a major concern in the clinical benefits sector. The crucial information is sourced from reliable sources and private protection groups. AI and data analytics are mostly utilized for disclosing healthcare fraud. There is no one technique or model that can be used to conceal all cases of clinical consideration blackmail. It will generally be concluded from this study that the general AI methodology, Sentiment Data Analysis and recently secured sources of clinical consideration data will be pivotal issues needed to make clinical consideration sensible, address the deficiency of clinical consideration distortion area, and provide top quality on clinical benefits systems.

References

1. B. Liu, Handbook Chapter: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing," Handbook of Natural Language Processing. Marcel Dekker, Inc. New York, NY, USA, 2009.
2. K. Dave, S. Lawrence and D.M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in Proceedings of the 12th international conference on World Wide Web, 2003, pp. 519–528.
3. K. Reynolds A. Kontostathis and L. Edwards, using machine learning to detect cyberbullying. In: Proceedings of the 10th international conference on machine learning and applications, 2011, 241–244.
4. Yue, Lin, et al. A survey of sentiment analysis in social media. Knowledge and Information Systems, 2018, 1-47.
5. D. Saurabh and Nitin N. Pise. Sentiment Analysis Methods and Approach: Survey. International Journal of Innovative Computer Science & Engineering 4.6, 2017, 7-11
6. M.N. Sarkies, K.A. Bowles, E.H. Skinner, D. Mitchell, R. Haas, M. Ho, K. Salter, K. May, D. Markham, L. O'Brien, S. Plumb, and T.P. Haines, Data Collection Methods in Health Services Research, Appl. Clin. Inform., 2015, 6(1), 96-109.
7. L. Qi and M. Vasarhelyi, Healthcare fraud detection: A survey and a clustering model

incorporating Geo-location information. In 29th World Continuous Auditing and Reporting Symposium (29WCARS), Brisbane, Australia. 2013.

8. A.R. Bauder and M.T. Khoshgoftaar, 15th IEEE International Conference on Machine Learning and Applications (ICMLA) - A Probabilistic Programming Approach for Outlier Detection in Healthcare Claims. 2016, 347–354.
9. S. Altuntas, T. Dereli and Z. Erdoğan, Evaluation of service quality using SERVQUAL scale and machine learning algorithms: a case study in healthcare, *Kybernetes*, 2022, 51(2), 846-875.
10. J.B.B. Rao, M.J. Rao, M. S. Rao, A.D.S. Saketh and T.R. Kumar, Optimizing the design of a fly wheel using machine Learning. Neuroquantology, 20(12), 533-254, 2022.
11. J.B.B. Rao, M.J. Rao, S. Paparao, A.D.S. Saketh and P. Anjaneyulu, computational analysis of a knuckle jointand implementation of the generalized regression neural network, 20(12), 3260-3271|, 2022.
12. M.J. Rao, M. Divya, M. R. Mohitha, P. Prasanthi, S. Paparao and M. Ramanaiah, Analyzing the Effectiveness of Convolutional Neural Networks and Recurrent Neural Networks for Recognizing Facial Expression. Int. J. Food and Nut. Sci., 11(7), 1269-1282, 2022**.**
13. M.J. Rao, P. Prasanthi, P.S. Patnaik, M. Divya, J. Sureshkumar and M. Ramanaiah, Forecasting Systems for Heart Disease Using Advanced Machine Learning Algorithms. Int. J. Food and Nut. Sci., 11(7), 1257-1268, 2022.
14. M. Balakrishna, M. Ramanaiah, B. Ramakrishna, M.J. Rao and R. Neeraja: Inductively Coupled Plasma-Mass Spectroscopy: Machine Learning Screening Technique for Trace Elemental Concentrations in Hemidesmus Indicus. *Annals of Forest Research*, November, 2022, 65(1): 4431-4445.
15. S. Jawale and S.D. Sawarkar, Interpretable Sentiment Analysis based on Deep Learning: An overview, 2020 IEEE Pune Section International Conference (Pune Con) Vishwakarma Institute of Technology, Pune India. Dec 16-18, 2020.
16. M.J. Rao, R.K. Kumar and J. Harikiran, Method for follicle detection and ovarian classification in digital ultrasound images using geometrical features. Journal of Advanced Research in Dynamical and Control Systems, 11, 1249-1258, 2019.
17. M.J. Rao and R.K. Kumar, Follicle Detection in Digital Ultrasound Images Using BEMD and Adaptive Clustering Algorithms, Lecture Notes in Mechanical Engineering. 2020, pp. 651–659.