IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

ADVANCED MACHINE LEARNING FOR MEDICAL IMAGE MINING: FUTURE OPPORTUNITIES

V V Nagaraju Goriparthi[1], K.L.V.G.Krishna Murthy[2], Gogineni Rajesh Chandra[3]

Department of EEE, Acharya Nagarjuna University, Guntur, AP, India
Associate Professor, Department of CSE, Vignan's Lara Institute of Technology & Science, Guntur, AP, India.
Professor, Department of CSE, KKR & KSR Institute of technology and Sciences, Vinjanampadu, Guntur, AP, India
nagaraju006@gmail.com, vngplkrishnamurthy@gmail.com[1], grajeshchandra@gmail.com

ABSTRACT

In medical image mining broadly used machine learning algorithms for effective classification and clustering.Due to the medical image mining to develop public health and eminence of life. The field of machine learning is rising the medical image analysis. All medical image applications connect to machine learning techniques. Medical image analysis denotes to set of measures to get valuable and correct information from multiple image processing modules for diagnosis.The proper extracted information helps improve the diagnosis according to the requirements of patients.This paper used a few well-known machine learning algorithms like k-nearest neighbors, support vector machine, Logistic Regression, and Random Forest tree on medical image data. Statistical results show the efficiency of algorithms on numerous metrics.

Keywords: Machine Learning algorithms, Medical Image Mining, Healthcare, Metrics.

1. INTRODUCTION

Data Mining is a process of extracting the images from the databases. It is also used for recollecting deleted images. Enhance the weak images with quality. Image retrieval is possible from databases and object recognition.

1.1 Need for Image Mining

Image mining is the most important field of medical scenarios and extract connections and patterns from raw data.It is a crucial technique based on Artificial Intelligence, Data Mining, Deep Learning, and Machine Learning.Numerous fields used image mining capabilities to discover novel image patterns.Image mining deals with a huge collection of images associated with alphanumeric data. The importance of image mining capability is to infer knowledge from data mechanically.It also handles raw images like the low-pixel illustrationsare managed effectively and efficiently to extract relationships of high level of objects. But up to now image mining is still in the investigational stage. Image mining concepts extract useful knowledge from image patterns.

Data mining and machine learning techniques are hugely demanding with its unique features for image data handling. Image mining affords a special framework that is used for raw

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

images in the database, which can't be used directly. Initially must be processed high-level modeling to use these techniques.Retrieving the patterns and knowledge from a large number of images for user interaction purposes image mining method is considered as optimal method [2] database.

Medical image analysis mainly focuses on convertingraw images into meaningful digitalized images for processing easily. Medical image research is more useful for the healthcare industry. It also creates challengesfor image quantity in complex and not solved issues. In real-timelarge-scale medical database processing purposes use some techniques based on user interest and domain knowledge. Generally, medical doctors require such type of simplified databases.At the time of implementation of the analysis so many restrictions impacted on effectiveness of image analysis. Daily large amounts of medical images are stored digitally. The following diagram figure 1 shows the life cycle of machine learning.

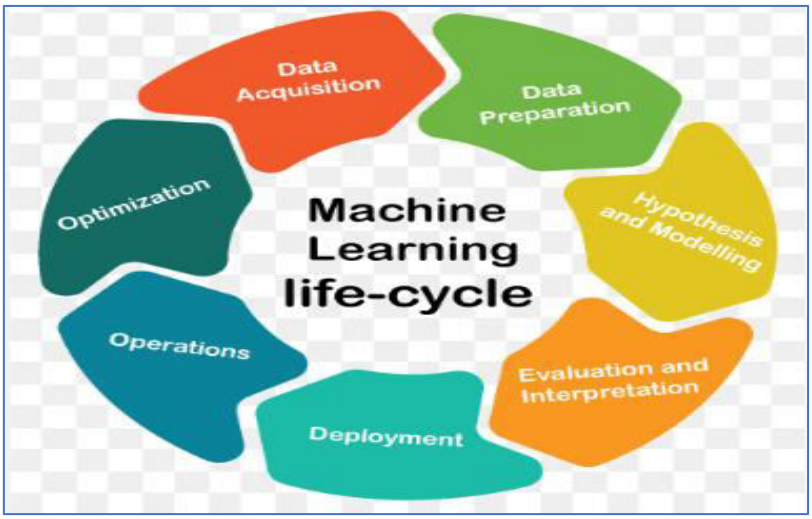

Figure 1: The Life Cycle of Machine Learning [1]

Our research describes the optimal methods of machine learning for medical image mining. These research results are based on different datasets, algorithms, model building, and various parameters. Finally, talks about medical image mining based on different factors.Currentlyour paper focuses on machine learning techniques only. Other than machine learning artificial intelligence and deep learning techniques are also available.

The following paper, next section discusses the background work of medical image mining. Section three states the medical image analysis. Section four discusses about Machine learning architecture of the procedure. Section five describes the results and analysis of multiple machine learning algorithms. The final section concludes the paper.

2. BACKGROUND WORK

Our literature survey focuses on machine learning techniques papers and medical image data papers. Medical image analysis mainly focuses on converting raw images into meaningful digitalized images for processing easily.Image mining affords a special framework that is used for raw images in the database, which can't be used directly. Initially must be processed high-level modeling to use these techniques.a few well-known machine learning algorithms

like k-nearest neighbors, support vector machine, Logistic Regression, and Random Forest tree on medical image data.

Initially,data mining techniques were used for image mining for analysis of data. These data mining techniques are used in large applications like educational organizations, the Insurance industry, medical diagnosis, space research, agricultural field, and industrial work. Image mining extracts the existing patterns and knowledge extraction but it is not suitable for storage databases.However advanced technology support for data extraction and storage capability and it supportsthe development of large image mining analysis and storage [3].

Data mining procedures are successful in image mining after that machine machine-learning techniques are used for the diagnosis and risk calculation. Machine learning techniques also face so many challenges in medical image mining. Issues like image protocol variations, weak labels, rendering, and assessing results [4].

The analysis of medical images is processed in multiple ways for learning models optimally,hierarchically, and rule-based algorithms. Large-label data also creates issues for both machine learning and conventional algorithms.Sometimestraining database is more complex compared to feature extraction of machine learning techniques [5]. Many practical applications provide different types of algorithms for current and future trends in image and text mining [10].

3.MEDICAL IMAGE MINING

Medical image mining focuses on low-level pixel analysis and they also construct mathematical models and rule-based structures to solve specific jobs.In the early 1990s, supervised machine-learning techniques were most popular for medical image analysis [6]. Innovation techniques introduced by pattern recognition and machine learning methods for image analysis accurately.

Machine learning techniques include computational data models for solving complex research issues [7].Discuss two types of image mining techniques in machine learning: supervised and unsupervised. In the machine learning process training examplesplay a crucial role in testing data to deliver optimal output. These techniques predict the future trends of results [8]. These algorithms default include the mathematical model, which can associate input data with labels and predict the unseen test result.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

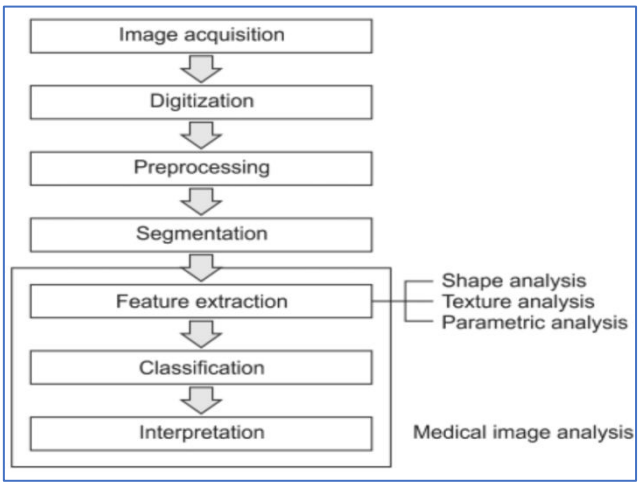

Figure 2: Hierarchy of medical image analysis

The following Figure 2 describes the hierarchy of medical image mining. The starting stage if mining is accepting the input which means the image and then converted into digitalization. After that perform the pre-processing of the given input. The next stage is segmentation and feature extraction (shape analysis, Texture analysis, and parametric analysis). The next stage is classification of data based on feature extraction and finally interpretation.

4. ARCHITECTURE OF MACHINE LEARNING

Machine learning systems consist of more components in the working organizational structure. It deals with the starting stage to the end stage of machine learning. Multiple stages are included in this procedure from pre-processing the data, feature selection, selection of algorithm, building the model, training,evaluatingthe models, and finally predicting the results.Every machine learning application based on the requirement of the user, model training, selection of datasets, size of the data, feature selection, and algorithm selection. The following figure 3 shows the process of machine learning architecture.

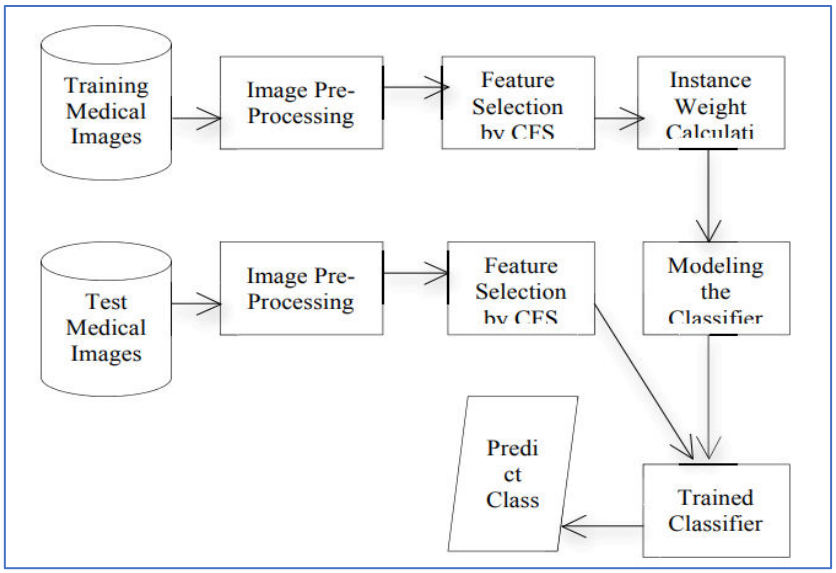

Figure 3: Proposed architecture of medical image classification

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

The following diagram 3 shows the architecture of medical image classification with different stages starting from the input of the medical images.Then pre-process the input datato remove noise and disturbances. Feature selection by CES method and then calculating the weights. Build the model and train the classifier data from feature selection. After completion of training test the medical images. Finally, predict the results.

4. RESULTS AND ANALYSIS

A. Random Forest Tree Model

This method trained with 150 trees with the highest depth of 40. For the measuring quality of splits used cross entropy method. In this process image segmentation job createsmultiple levels of complexity [11]. The following series run in a random forest tree is g(x)=f0(x)+f1(x)+f2(x)+.....

B. Support Vector Machine

The support vector machine was trained with 300 images with the optimal hyperplanemaking predictions based on the learning model. The task of SVM use the appropriate kernel functions with accurate constraints [12]. To find binary class classificationN-dimensional space. The following formula is $\forall n : | y n - (x n ' \beta + b) | \leq \varepsilon$.

C. K Nearest Neighbour

This method calculates every training instance based on the similarity with every test instance.Generally,in this process select the neighbourhood size is 5-10% of the training set.KNN method took input as the training set as medical images and generated each image weight of the testing set [13].The most common choice is the Minkowski distance dist(x,z)=(d∑r=1|xr−zr|p)1/p.

D. Logistic Regression

Logistic regression of the Machine learning model is used for splitting conditions and tested test data for predictions to get optimal accuracy and discover behaviour of model. The algorithm predictions show 1 and 0 for positive or negative [14].Logistic regression and its function. (5.2) $S (x) = 1 2 [1 + \tanh x 2]$, $\tanh x = e x - x - x e x + e - x$.

The following two tables show the hardware and time requirements of model training and performance comparison with multiple metrics.

Table 1: Systemhardware and time requirements for model training

Algorithm		**Memory(Approximately)**	**Time**
Random Forest Tree	Feature Extraction	207 GB	7 Hrs
Support Vector Machine			7.3Hrs
KNN			7.5 Hrs
Logistic Regression			7.9 Hrs

Table 2: Performance comparison through multiple measurements

Algorithm	**PhC-C2DH-U373**				**Retinal Images**			
	Accuracy	**Dice coefficie**	**IoU**	**Sensitivity**	**Accuracy**	**Dice Coefficie**	**IoU**	**Sensitivity**

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

		nt				nt		
Random Forest Tree	0.98	0.85	0.75	0.84	0.95	0.68	0.52	0.58
Support Vector Machine	0.93	0.82	0.73	0.81	0.91	0.65	0.51	0.56
KNN	0.90	0.80	0.71	0.79	0.88	0.63	0.48	0.53
Logistic Regression	0.89	0.78	0.69	0.76	0.86	0.62	0.45	0.51

In this research,every algorithm data can be split into train and test data 80% and 20% respectively. All machine learning algorithms follow the same procedure for analysis of data, finally predict the results. The hardware and time requirements are described in Table 1. Performance comparison of algorithms is shown in Table 2. Figure 4 discusses the predictions of the dissimilar algorithms compared to one image of every dataset ofPhC-C2DH-U373 and Retinal Images. Figure 5 states the performance measured against various evaluation metrics in comparison using boxplots.

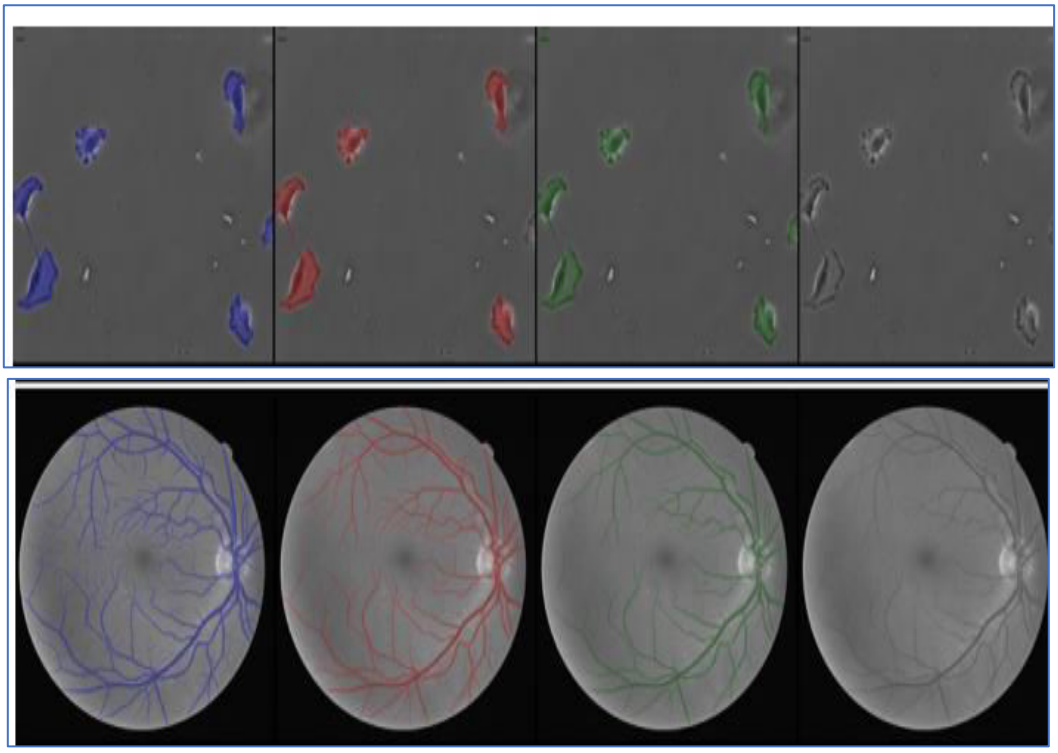

Figure 4: Predictions of the dissimilar algorithms compared to the one image of everydataset. From the table 1 and 2 describe the information of multiple algorithms with two datasets. Table 1 discusses hardware and time necessities for model training and implementation. Table 2 states the performance comparison through multiple measurements of four algorithms.Four measurements were calculated in this experiment. They are accuracy, Dice, IoU, and Sensitivity.

Accuracy = (TP + TN) / (TP + TN + FP + FN) (1)
Sensitivity= TP / (TP + FN) (2)
IoU= TP / (TP + FP +FN) (3)
Dice = 2 * (No. of Common elements) / (Total no. of elements) (4)

The accuracy metric is used tocorrectly predict the outcome. Dice metric describes the model and debugs their models for potential problems.IoUis a recital metric used to assess the accuracy of footnote, dissection, and object discovery algorithms.Sensitivity refers the how machine learning model detects the positivity instances. Among the four algorithms, maximum accuracy is predicted by a random forest tree with 98% of the remaining three algorithms Support Vector Machine, K-Nearest Neighbour, and Logistic Regression are 93%, 90%, and 89% respectively.

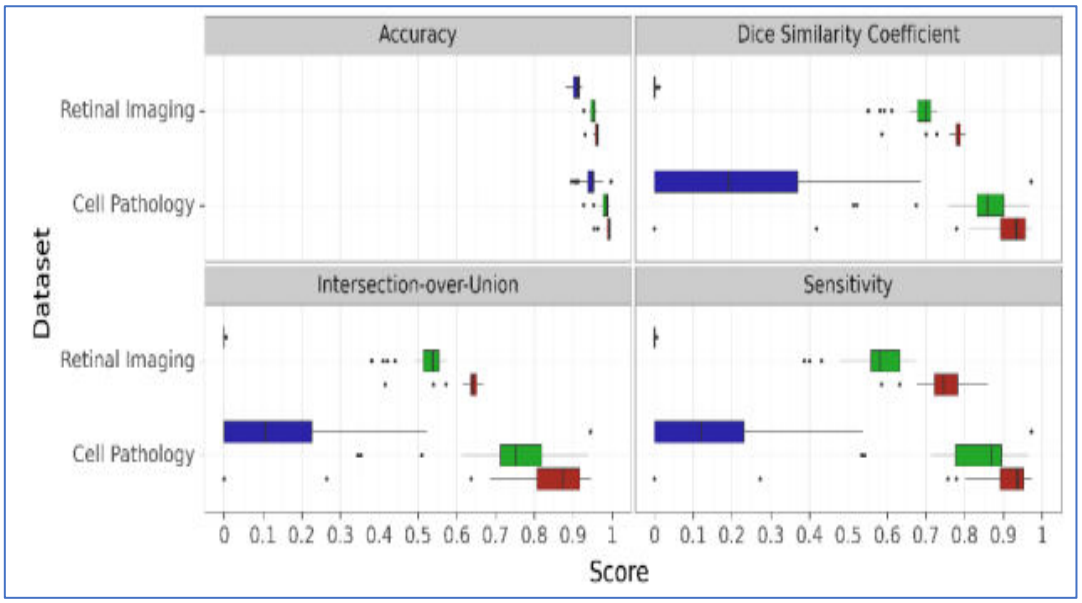

Figure 5: Algorithm recitals for the datasets via boxplots of segmentation

Our evaluation shows the accuracy, dice, IoU, and sensitivity values of four algorithm models predict multiple results. Less memory requirement for algorithm implementation. Feature extraction from pre-processed data reaches accurate predictions. After analysis of images retrieve the effective acknowledgment of a piece of an image and enhance the ability to identify patterns of images. Among four models Random Forest allows more flexible access to detailed medical image dissection [15]. One drawback of this algorithm for reaching optimal results is requiresa large amount of hardware memory.Numerous studies describe the best outcome for random forest tree medical images compared to the remaining three models.

5. CONCLUSION

The field of machine learning is rising the medical image analysis. All medical image applications connect to machine learning techniques. Medical image analysis denotes to set of measures to get valuable and correct information from multiple image processing modules for diagnosis. The proper extracted information helps improve the diagnosis according to the

requirements of patients.In addition, some common algorithms were applied like k-nearest neighbors, support vector machine, Logistic Regression, and Random Forest on medical datasets to check the efficiency of algorithms on different metrics. Among the four algorithms, maximum accuracy is predicted by a random forest tree with 98% of the remaining three algorithms SVM, KNN, and LR are 93%, 90%, and 89% respectively.

5.1 Future Enhancement

The future enhancement of medical image mining by applying deep learning techniques with CAD support for clinical and enhancing the accuracy and efficiency of different diagnoses. Artificial intelligence medical imaging also improves the accuracy and efficiency of disease diagnosis[16] and these techniques assist the healthcare professional in detecting abnormalities and finding the structure of diseases.

References

1. Available at: https://techblog.cdiscount.com/a-brief-overview-of-automatic-machine-learning-solutions-automl/
2. Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R. and Virage CFS (1996) 'Image search engine: an open framework for image management', Storage and Retrieval for Image and Video Databases, SPIE, Vol. 6, No. 76, pp.76–87.
3. Nilanjan Dey et. al," Image mining framework and techniques: a review", Int. J. Image Mining, Vol. 1, No. 1, 2015.
4. Marleen de Bruijne," Machine learning approaches in medical image analysis: From detection to diagnosis", Medical Image Analysis 33 (2016) 94–97.
5. A. Criminisi," Machine learning for medical images analysis", Medical Image Analysis 33 (2016) 91–93.
6. William M. Wells III," Medical Image Analysis –past, present, and Future", Medical Image Analysis 33 (2016) 4–6.
7. Hagan MT, Demuth HB, Beale MH, De Jesus O. Neural network design. 2nd ed. Boston (MA): PWS Publishing; 2014.
8. Le QV, Ngiam J, Coates A, Lahiri A, Prochnow B, Ng AY. On optimization methods for deep learning. In:Getoor L, Scheffer T, editors. Proceedings of the 28th International Conference on Machine Learning; 2011 Jun 28- Jul 2; Bellevue, WA. Madison (WI): Omnipress; 2011. p. 265–72.
9. Dr. R. Vijaya Kumar Reddy, Dr. Shaik Subhani, Dr. G. Rajesh Chandra, Dr. B. Srinivasa Rao," Breast Cancer Prediction using Classification Techniques", International Journal of Emerging Trends in Engineering Research, Vol. 8, No.9,2020.
10. Mr. Sujan Reddy, Ms. Renu Sri, and Subhani Shaik," Sentimental Analysis using Logistic Regression", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.
11. R. Vijaya Kumar Reddy, Subhani Shaik,B. Srinivasa Rao, "Machine learning based outlier detection for medical data" Indonesian Journal of Electrical Engineering and Computer Science, Vol. 24, No. 1, October 2021.
12. Subhani shaik, Vinay Kumar," Patient expenditure prediction in health care sector", International Journal of Science and Research, Vol. 11, Issue 6, June 2022.
13. Subhani Shaik, "Pattern Recognition of Speech Signals Using Curvelet Transform and

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 06, 2022

Artificial Intelligence", International Journal of Creative Research Thoughts, Vol. 6, Issue 2, April 2018.

14. Md. Samiuddina, G.Rajender, K. Sai Abhiram Varma, A. Ravi Kumar, and Subhani Shaik,"Health Insurance Cost Prediction Using Deep Neural Network",Asian Journal of Research in Computer Science, Volume 16, Issue 2, Page 46-53, June- 2023.
15. P. Pranathi, V. Revathi, P. Varshitha, Subhani Shaik and Sunil Bhutada,"Logistic Regression Based Cyber Harassment Identification", Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 76-85, June-2023.
16. Bopparaju Supriya, Jajula Om Sri Keerthi, PeekaUshasvi Rachel, Usha Manjari, Subhani Shaik,"Image Fusion with CT and MRI Images for Improving the Quality of Diagnostics", Accelerating discoveries in Datascience and Artificial Intelligence I, ICDSAI 2023, LietVizaianagaram, April 24-25, India. https://link.springer.com/chapter/10.1007/978-3-031-51167-7_39