

Data Visualization for the Prediction of Liver Cancer Disease using Different Graphical Techniques



Gururaj L Kulkarni, Sanjeev S Sannakki, Vijay S Rajpurohit

Abstract: Data visualization is the technique for analyzing the data from the collected dataset. Different plots can be drawn for the data visualization. Microscopic images of the liver are being collected as a dataset from the authorized laboratory and the Joint plot, Violin plot and distribution plot are applied on them for the analysis which helps to extract the specific features and for the classification. Joint plot uses the scatter plot and Histogram technique in order to visualize the data. Violin plot technique is used for plotting the numeric data which helps in gray level co-occurrence matrix. Distribution graph is plotted to check the distribution of tones captured in the image so that we can differentiate based on the tones. All three graphs plotted extract the different features which help in efficient analysis.

Keywords: Dataset, Data visualization, Joint plot, Violin plot, Distribution plot.

I. INTRODUCTION

One of the important organs of our body is liver as it carries out vital roles in our body such as drugs metabolism, toxin clearance and blood protein production. Any damage to the liver would lead to serious complications in its functionality which would further lead to liver failure. Damage to the liver can happen due to several reasons such as alcoholism, obesity, viral hepatitis etc.; some disease would cause serious complications which may further lead to liver transplant. If the disease related to the liver is identified at prior stage then there are chances of saving the liver from serious problems. There are many medical imaging modalities that help us to diagnose the diseases related to any organs. The different scanning techniques are used by the doctors to understand the disease related to the liver and for further diagnosis [7]. This process actually involves a radiologist who will do the scanning of the patients liver, gives details regarding the structure of the liver by thorough observation and then the forwards the image to the doctor, then doctor makes the diagnosis of the liver by his knowledge as well as the details

given by the pathologist[8]. One of the important characteristics of image is texture, which can be seen in all images. The texture of image is having varying pixel intensities i.e. gray values analyzing pixel intensities is useful for wide variety of applications.

II. METHODOLOGY:

The dataset is being collected from the JJM medical college and the following steps are applied for the visualization of the data.

- Step 1: Collect the different combination of the dataset.
- Step 2: For numeric analyses of the data go to step 5.
- Step 3: For distributions of variables in the data go to step 6.
- Step 4: For statistical relationship go to step 7.
- Step 5: Plot violin graph.
- Step 6: Plot distribution graph.
- Step 7: Plot joint graph.

III. LITERATURE REVIEW

Ahmadian A et al. proposed a method for categorizing different liver diseases using Gabor wavelet texture feature extraction method and classified ultrasound liver images into normal type, cirrhotic type and hepatitis type. Dyadic wavelet transform as well as Gabor wavelet have been applied to extract the features from the images. An accuracy of around 87% was observed in the classification [1].

Balasubramanian D et al. came up with an automatic classifying system which classified liver into benign category, malignant category, cyst category and normal liver category. Around 4 features extraction methods were applied namely SGLDM, RUNL, TEM and Gabor wavelets techniques. K-means clustering algorithm and BPN algorithm were applied for classification. K-means gave result around 80.75% and BPN 84.6% respectively. Performance of BPN is better comparatively than K-means algorithm [2].

Hwanga Y N et al. focused on the development of a diagnostic system to accurately classify different focal liver lesions. Around 29 images of cyst category, 37 images of hemangiomas and 33 images of malignant category were used for the study. Ultrasound images were used for the analysis. Different feature extraction methods were utilized for the work. Features were composed of these methods first order statistics; gray level co-occurrence matrices, Law's. Some features were acquired from Principal component method. Feed-forward neural network was used for the classification.

Revised Manuscript Received on April 25, 2020.

* Correspondence Author

Gururaj Kulkarni*, Assistant Professor, Department of Information Science and Engineering, Gogte Institute of Technology, Belgaum, Karnataka, India.

Dr. Sanjeev S Sannakki, Professor, Department of Computer Science and Engineering, Gogte Institute of Technology, Belgaum.

Dr. Vijay S Rajpurohit, Professor, Department of Computer Science and Engg. KLS Gogte Institute of Technology, Belagavi, Karnataka, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

It gained an accuracy of around 96%. Echogenicity feature showed good improvement in the overall result [3]. Neogi N et al. analyzed a system for classifying the diseases related to liver. The basic idea in the proposed system was to classify the liver into normal liver category and fatty liver category. For this eleven statistical textural features extracted from US images were used. For this study around 14 fatty images of liver and 28 images of normal livers were used. The study used unsupervised and supervised neural network for the purpose of classification [4].

IV. EXPLORING DATASET

The dataset is being collected from the JJM Medical college and data visualization is being done on this dataset for the efficient analysis.

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...
2	84300903	M	19.69	21.25	130.00	1203.0	0.10360	0.15990	0.1974	0.12790	...
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...
4	84388402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...

5 rows × 33 columns

Figure 1: Dataset

V. RESULTS

Joint Plot

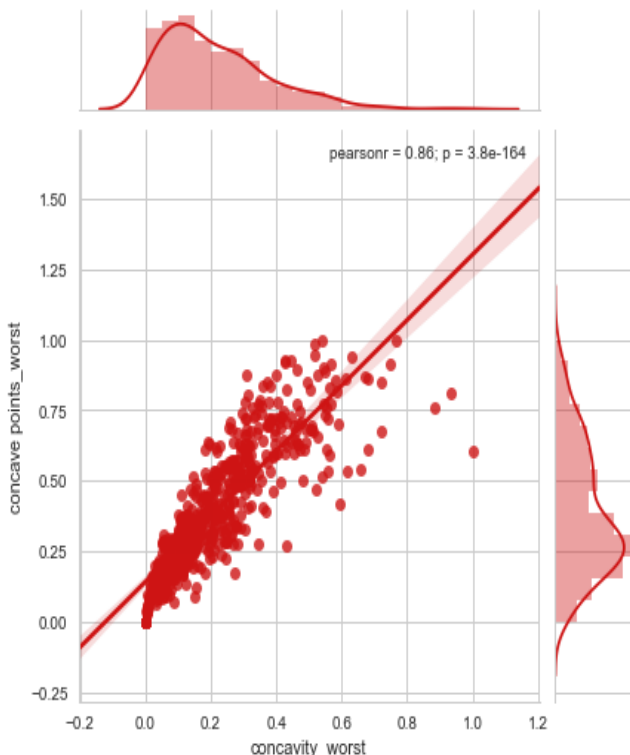


Figure 2: Joint Plot

The data visualization is being done on the dataset by plotting joint plot graph. Figure 2 shows the joint plot which uses the scatter plot and Histogram technique in order to visualize the data.

Violin plot

A violin plot technique is used for plotting the numeric data. Figure 3 shows the violin plot as a combination of box plot and kernel density plot. In figure 3

- The '-' on the violin plot represents the median.
- The black bar in the centre of the violin represents the interquartile range.
- The black lines stretched from the bar represents the lower/upper adjacent values.

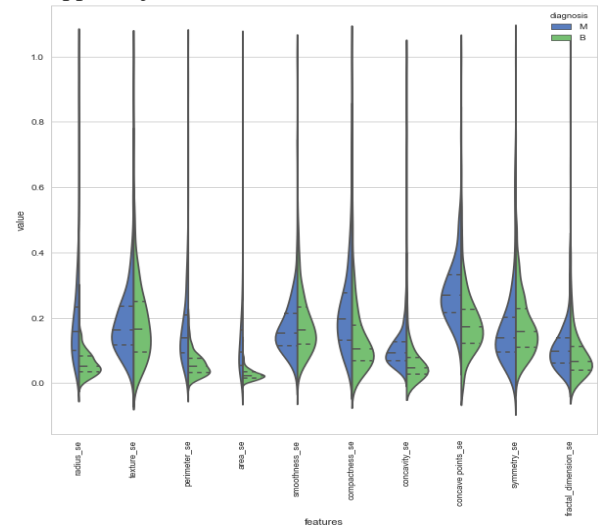


Figure 3: Violin plot

Distribution Plot

Distribution plot can be used to analyze how the variables are being distributed in a dataset. Figure 4 shows the distribution plot on the collected dataset.

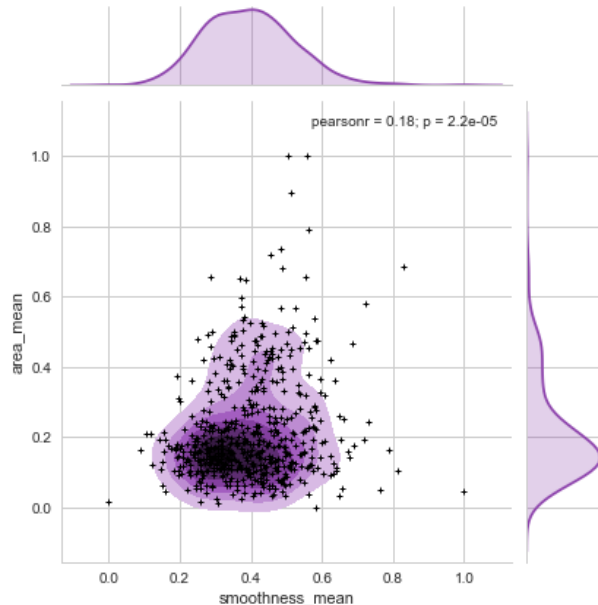


Figure 4: Distribution plot

VI. CONCLUSION:

The data visualization is being done by plotting Joint plot, Violin plot and distribution plot on the collected dataset. The parameters contributing to Liver cancer were scrutinized using the different visualization techniques. It helps to analyze the data in a more efficient manner.

REFERENCES

1. Ahmadian A, Mostafa A, Abolhassani M.D, and Salimpour Y, "A texture classification method for diffused liver diseases using Gabor wavelets", 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005- Shanghai- China, pp. 1567 – 1570, 2005, Print ISBN: 0-7803-8741-4.
2. Balasubramanian D, Srinivasan P and Gurupatham R, "Automatic classification of focal lesions in ultrasound liver images using Principal Component Analysis and Neural Networks", 29th Annual International Conference of the IEEE Engineering and Medicine and Biology Society- 2007, Lyon-France, pp.2134-2137, 2007, Print ISBN: 978-1-4244-0787-3.
3. Hwanga Y N, Lee JH, Kim GY, Jiang YY and Kim SM "Classification of focal liver lesions on ultrasound images by extracting hybrid textural features and using an artificial neural network", Bio-Medical Materials and Engineering, Shanghai-China , pp. S1599–S1611, 2015.
4. Neogi N, Adhikari A, Roy M "Classification of Ultrasonography Images of Human Fatty and Normal Livers using GLCM Textural Features", Current Trends in Technology and Science, Vol 3, pp.252-259, 2014.
5. Gururaj L Kulkarni, Sanjeev S Sannakki, Pavan N Kunchur, Vijay S Rajpurohit, "Data Analysis for the prediction of the liver cancer disease using features of the dataset", International Journal of Engineering, Applied and Management Sciences Paradigms (IJEAM) , ICRMR-2019 , ISSN 2320-6608 ,pp 57-60.
6. Reetu and Narendra Kumar, "Creative Commons Attribution CC BY Medical Diagnosis for Liver Cancer using Classification Techniques" 2013. International Journal of Science and Research (IJSR) ISSN, Volume 4, Issue 5.
7. Urban, Jan & Vanek, Jan & Stys, Dalibor, "Preprocessing of microscopy images via Shannon's entropy", Conference on Systems Biology of Mammalian Cells, At Freiburg, Germany, pp 434-438.
8. Rajamani.R , Rathika "Analysis of Liver Cancer using Adaptive Neuro Fuzzy Inference System (ANFIS)" July 2015 , International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 7.
9. Alexander, Joshua, Mahmood, John and Ray "A signature based liver cancer predictive system" Information Technology: Coding and Computing, International conference on (2005) April 4 -6 2005, pp 195-199.
10. Kemal Polat,Seral, Seralsahan, Halife Kodaz, Salih Gunes,"Breast cancer and liver disorders classification using artificial immune recognition system (AIRS)" Elsevier pp172-183.
11. Rajamani.R , Rathika "Analysis of Liver Cancer using Adaptive Neuro Fuzzy Inference System (ANFIS)" July 2015 , International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 7.
12. P. Rajeswari and Sophia "Human Liver Cancer Classification using Microarray Gene Expression Data" November 2011, International Journal of Computer Applications (0975 – 8887) Volume 34– No.6.
13. Reetu and Narendra Kumar, "Creative Commons Attribution CC BY Medical Diagnosis for Liver Cancer using Classification Techniques" 2013. International Journal of Science and Research (IJSR) ISSN, Volume 4, Issue 5.
14. SS Sannakki, VS Rajpurohit, "An approach for detection and classification of leaf spot diseases affecting pomegranate crop", International Journal of Advance Foundation and Research in Computer, vol 2, 2015.
15. Anisha, Kishor kumar Reddy, Narsimha prasad "A pragmatic approach for detecting liver cancer using image processing and data mining techniques" IEEE 2015.

AUTHORS PROFILE



Gururaj L Kulkarni is working as an Assistant Professor in the Department of Information Science and Engineering at Gogte Institute of Technology, Belgaum, Karnataka, India. He pursued his B.E. and M.Tech in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi. He is currently pursuing Ph.D. at the Visvesvaraya Technological University, Belagavi, India in the Department of Computer Science and Engineering. He has published a good number of papers in journals, International, and National conferences. His main research interests include Image processing, Machine Learning .



Dr. Sanjeev S Sannakki has completed his Ph.D. degree in Image processing & Data Mining from VTU Belagavi. His career spans over a period of two decades in the field of teaching, research and other diversified in-depth experience in academics. He is currently working as a Professor in the Department of Computer Science and Engineering, Gogte Institute of Technology, Belgaum. Currently, he is shouldering the responsibility of Head of the Research center. He has published several papers in reputed national/international conferences and journals. He is also guiding the research scholars & UG/PG students of VTU.



Dr. Vijay S Rajpurohit is working as Professor, in the department of Computer Science and Engg at KLS Gogte Institute of Technology , Belagavi, Karnataka, India. He did his B.E. in Computer Science and Engg from Karnataka University Dharwad, M.Tech at N.I.T.K Surathkal and PhD from Manipal University, Manipal. His research areas include Image Processing, Cloud Computing, Machine Learning and Data Analytics. He has published a good number of papers in journals, International and National conferences. He is the reviewer for few international journals and conferences. He is the associate editor for two international journals and Senior Member of International Association of CS and IT. He is also the life member of SSI, ISC and ISTE associations.