

R. Raja Mani¹,
Dr. Chandani Sharma²

Evaluation of Feature Selection Techniques for Optimal Attribute Identification in Image Processing Applications



Abstract: Texture analysis is very important for many image processing and pattern recognition applications, especially in biomedical diagnostics where small differences in tissue texture must be understood correctly. The main goal of this study is to create and test a modified noise-robust local binary pattern (MNRLBP) method. The main goal is to find and choose the best textured features that improve the speed of classification and speed of calculation, especially in hard imaging conditions that come when diagnosing liver problems using ultrasound. The approach includes pre-processing and generalization datasets such as outx_tC_00010, outx_tC_00012, broadhetz and a liver ultrasound dataset. This study used modified LBP to obtain an encoded descriptor locally, including both local signals and magnitude information for convenience extraction. Three strong selection methods-average-range plot, box plot analysis, and fissure decimant ratio (FDR) -Are is then used to improve features. This reduces the number of dimensions and increases the ability to explain the difference between them. Results suggest that the suggested MNRLBP approach increases classification performance by keeping the significant macro structural and microstractive information while being resistant to rotation and noise. Using each selection approach, the best was found from 53 recovered features. These multitude were better in separating common and cirrotic liver tissues into separate classes. These results suggest that the two work together to analyze medical images by using both integrated textures descriptor and statistical selection processes. This study has a major impact on clinical decision making as it gives doctors an accurate and automatic way to classify diseases. Technology can also be used in other areas, which require very accurate texture characterization such as remote sensing, material inspection and facial identification systems.

Keywords: Feature Selection; Optimal Attribute; Image Processing; Fisher Discriminant Ratio

Introduction:

The texture is an important visual feature that helps us understand and classify images, especially when it comes to natural and biomedical surfaces. To find and explain the difference between normal and sick tissues, medical imaging, especially in liver ultrasound diagnostics is very important (Nahed et al., 2020; Bolon-Canado and Remissy, 2020). Like other biological surfaces, the liver texture includes complex visual patterns made from intensity, shape and structure changes, making them difficult to classify accurately. Traditional methods of medical experts to explain images are sometimes influenced by subjective and noise, light and operator experience (Remissy, and Bolon-Canado, 2019; Venkatesh, and Anuradha, 2019).

Computer methods that automatically find and choose important texture features have become more and more important to achieve these problems. Local binary patterns (LBPs) have become a popular and effective way because they are easy to calculate to occupy local texture features and do a great work. But traditional LBP methods are often sensitive to noise and macro cannot show structural information. This paper suggests a modified, noise-resistant LBP (MNRLBP) that connects both local and regional statistical fluctuations in the texture representation to achieve these problems. The method adds spatial information at the neighborhood and makes rotation invariance strong, which makes the texture classification more accurate. Choosing the most useful features is as important as extracting features to reduce the dyspation and increase classification performance (Casego and Sun, 2020; Remisiro, and Bolon-Canado, 2019). So, this study also sees several ways to choose features such as the best feature, the mining plot, box plot analysis and fisher discipline ratio (FDR). The results

¹ Department Of Electronics And Communication Engineering

^{1,2}Shri Venkateshwara University, Gajraula (Uttar Pradesh)

of the study have the ability to improve automatic liver disease diagnosis and other image-based classification tools. The following section explains the previous literature related to this study in detail.

Literature Review:

Recent research has shown how important feature selection is to improve the performance of machine learning models in many areas. Elaberry et al. (2022) showed that the use of the genetic algorithm to detect credit card fraud to detect the number of dimensions in the data and to get rid of the symptoms was more accurately accurately accurate that were not useful. Marcos-Jambrano et al. (2021) also saw how machine learning is used in human microbiom test. He focused on how the feature selection approaches help to find a biomarker, predict diseases and make personal treatment plans. Disha and Waheed (2022) came with a gini impurity-based weighted random forest (GIWRF) method for intrusion systems. He showed that this method of choosing features made the classification more accurate and cut on false alarms. Ghosh et al. (2021) also showed how important the good feature selection is to predict cardiovascular disease and methods of lasso. This made the model easier and more efficient to understand. These studies all show that the selection of the right features not only improves the accuracy of predictions, but also ensures that models are sharp and reliable in a wide range of applications..

Methodology:

The method of this study for texture-based image analysis uses a systematic way to extract, select and classify features with B-Mode Ultrasound Liver Tissue Examination such as medical imaging. First of all, processes such as noise reducing and normalization are taken to ensure that the photos are all similar. Advanced descriptors, including local binary pattern (LBP), Gabor filter, wavelet transforms, and co-phenomenon matrix-based algorithms, are used to achieve texture features as they both do a good job of capturing both fine and coarse textured information. It is also important to use rotation-invariant and multi-scale lbp variants to make the system more resistant for changes in image orientation and lighting. Structural combination matrices are then used to choose features that keep their discriminative power while lowering their dimensionality. This phase is very important to get rid of problems like noise interference and redundancy, which will make the processing faster and the model more accurate. In the classification step, machine learning algorithms use selected characteristics to sort photos into groups based on the texture patterns that are present in them. This technology not only helps with automated diagnosis, such finding liver problems, but it also makes sure that the results are objective and can be repeated, unlike visual assessments by doctors that are based on their own opinions.

Results And Discussions:

The findings of the feature selection phase using mean-range plots clearly show that the selected texture features can tell the difference between normal and cirrhotic liver tissues. The graphs in Figures 1 (a) and (b) show the normalized mean and range for the "Entropy" and "Inverse Moment Normalized" features, respectively. In the Entropy plot, the cirrhotic class always has a higher mean value and a smaller range than the normal class. This suggests that entropy is more stable and greater in cirrhotic areas, which means it is beneficial for classification. On the other hand, the Inverse Moment Normalized feature displays a lot of overlap between classes with comparable mean values and broad error bars. This means that it can't tell the difference between liver tissue conditions very well.

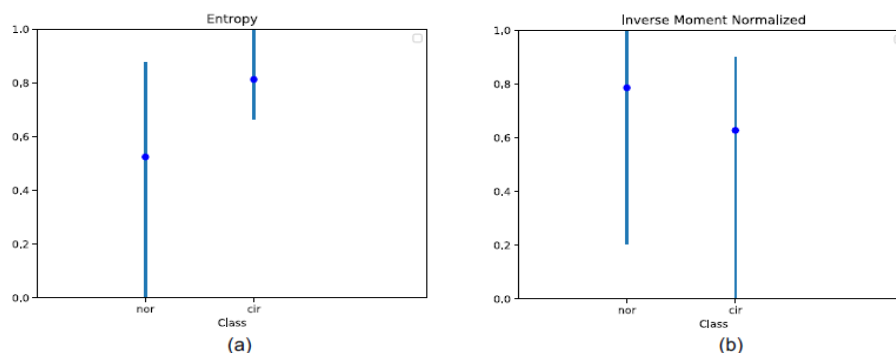


Figure 1: (a) Mean Range Plot showing selected feature, and (b) Mean Range Plot showing discarded feature

Also, the boxplots back up these results by showing how each feature is spread out statistically over both groups. The Entropy boxplot (Figure 2a) shows that the cirrhotic and normal classes are clearly separate with very little overlap, which supports keeping it as a key characteristic. The Inverse Moment Normalized boxplot (Figure 2b), on the other hand, shows a lot of overlap and outliers, which means that the classes don't separate very well. This is why it should be removed from the selecting process. The mean-range plots and boxplots together make a solid case for lowering dimensionality and focusing on textural qualities that can tell the difference between things the best. Normalization makes sure that comparisons are fair, and the method successfully extracts the features that are most likely to predict liver status, which improves the performance of the next classification model.

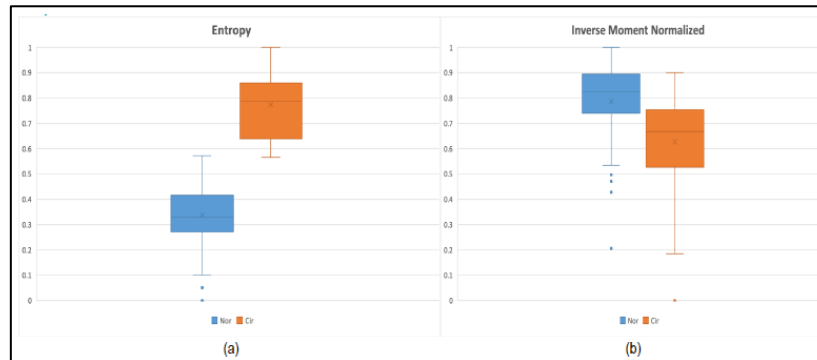


Figure 2: (a) Box Plot Showing selected feature, and (b) Box Plot showing discarded feature.

Conclusion:

The results of this study show that entropy-based texture features are very good at telling the difference between cirrhotic and normal liver tissues. Normalization and mean-range plots got rid of unimportant characteristics like inverse moment normalized, which improved the accuracy of the categorization. This shows how important it is to choose strong features when analysing medical images to get a solid diagnosis.

References:

- [1] Naheed, N., Shaheen, M., Khan, S. A., Alawairdhi, M., & Khan, M. A. (2020). Importance of features selection, attributes selection, challenges and future directions for medical imaging data: a review. *Computer Modeling in Engineering & Sciences*, 125(1), 314-344.
- [2] Bolón-Canedo, V., & Remeseiro, B. (2020). Feature selection in image analysis: a survey. *Artificial Intelligence Review*, 53(4), 2905-2931.
- [3] Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112, 103375.
- [4] Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybern. Inf. Technol.*, 19(1), 3-26.
- [5] Adebayo, M. Deepfakes and Data Privacy: Navigating The Risks in the Age of AI. NDPC-, 106.
- [6] Kasongo, S. M., & Sun, Y. (2020). Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. *Journal of Big Data*, 7(1), 105.
- [7] Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 9(1), 24.
- [8] Marcos-Zambrano, L. J., Karadzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovic, V., Aasmets, O., ... & Truu, J. (2021). Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in microbiology*, 12, 634511.
- [9] Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity*, 5(1), 1.
- [10] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, 9, 19304-19326.