Image Processing Strategies for Automatic Detection of Common Gastroenterological Diseases

Rafael Neujahr Copstein, Vicenzo Abichequer, Matheus Cruz Andrade, Lucas Almeida Machado, Evandro Rodrigues, Leonardo Pavanatto Soares, Marcio Sarroglia Pinho
Virtual Reality Group - School of Technology - PUCRS
Porto Alegre, Brazil
rafael.copstein@acad.pucrs.br, pinho@pucrs.br

***Abstract*—The analysis of Confocal Laser Endomicroscopy (CLE) is one of the techniques used for diagnosing gastroenterological diseases. However, the manual analysis of such images requires training and experience and will often lead to wrong diagnostics. This work explores the use of attributes taken from classic texture description techniques, gray level co-occurrence matrices (GLCM) and local binary patterns (LBP), as inputs for classifiers to separate images from 3 common gastroenterological diseases, with 262 images. A baseline classifier was trained for the 10 smaller groups and two others were trained using GLCM and LBP attributes. Overall, the benefits of using texture analysis techniques and attributes can be observed as an increase in accuracy and consistency of the results.**

Keywords—image processing; machine learning; esophageal cancer;

I. INTRODUCTION

Barrett's Esophagus and associated adenocarcinoma have emerged as a major health care problem over the last two decades [7]. It is also one of the major complications of gastroesophageal reflux disease commonly encountered in gastroenterology clinics [3].

One of the techniques used to diagnose the Barrett's Esophagus disease (and associated adenocarcinoma) involves capturing images from a Confocal Laser Endomicroscopy (CLE). The procedure helps to diagnose cancer in its premature stages, which can help preventing chemo and radiotherapy treatments, while also generating high quality images.

However, a pathologist is still required to analyze the images after they have been captured. This requires significant training and experience. According to Salomao et al [1], many diagnostics are downgraded by experts and many pathologists often over diagnose patients. This shows that this kind of diagnostic is still very subjective to the pathologist that gave it.

The need for a new technique to analyze CLE images comes from a need for giving more precise and non-subjective diagnostics to patients. In this paper, the use of different attributes for describing image textures as classifier inputs is explored, more specifically, the impact of gray level co-occurrence matrices and the use of local binary patterns in the classification process of common gastroenterological diseases is studied. The diseases explored here are: Gastric Metaplasia (GMP), Barrett's Esophagus (BAR) and Neoplastic Mucosa (NPL), also known as Esophageal Cancer.

For training and validating each of the classifiers, a dataset of CLE images of the three diseases was used. However, the dataset does not contain enough images to make deep learning techniques viable and it does not contain a balanced number of images for each category. The workaround for these issues is also explored.

This paper is organized as follows: section II gives some background on each of the diseases, section III analyzes the current state of related research, section IV introduces the dataset used in the experiment, section V present the developed methods and section VI discusses the results obtained in the experiments.

II. BACKGROUND

A. Gastric Metaplasia

Also known as Gastroesophageal Reflux Disease, the Gastric Metaplasia (GMP) is characterized by the return of content from the stomach to the esophagus due to a failure in the mechanisms that stop that kind of behavior [20]. This kind of reflux is, most of the time, composed of gastric acids that are not well handled by the mucosa (internal layer of the esophagus) and causes it to get inflamed. Figure 1 shows an example of a GMP image obtained using CLE.

B. Barrett's Esophagus

Also known as Intestinal Metaplasia, the Barrett's Esophagus (BAR) disease is characterized by the change in esophageal cells due to prolonged lesions [20]. This disease may be seen as an adaptation from the lesions caused by reflux since the changed cells are more resistant to the acid pH of the gastric reflux. It is characterized by a pink-colored region observed in the esophagus. Figure 1 shows an example of a BAR image obtained using CLE.

C. Neoplastic Mucosa

Also known as esophageal cancer, Neoplastic Mucosa (NPL) is a disease where malignant cells start developing inside the esophagus [20]. Due to the lesions caused by GMP and BAR it is very common to see patients with esophageal cancer having a history of contracting the other two diseases. Figure 1 shows an example of a NPL image obtained using CLE.

III. RELATED WORK

The usage of image analysis for classifying gastroenterological diseases has been explored by some published researches [2-9]. An analysis was conducted to assess the current state of research in the field.

Van der Sommen et al [2], describes a method that relies on machine learning to do the classification. First, the images are searched for regions of interest: the lumen, intestinal juices and specular reflections. Then, the images are split into 50x50 pixel blocks and the mean and variance of each block is calculated. They are used as inputs for a Support Vector Machine (SVM) which predicts whether a block shows neoplasia or not. The paper claims that the method presented gives results with 0.83 for both sensitivity and specificity on a per-image analysis. Given a patient's multiple exams, the system reached 0.86 and 0.87 of sensitivity and specificity, respectively.

In Dunn et al [3], another machine learning centric algorithm is presented for doing the classification. The paper aims to compare two techniques for classification: Nucleotyping (NT) and Image Cytometric DNA Analysis (ICM). NT uses a set of 35 attributes taken from each image's Gray Level Co-occurrence Matrix, Grey Level Run Length Matrix, and Grey Level Entropy Matrix. Most are attributes regarding the image's texture. ICM, on the other hand, was performed by an automatic image cytometric analyzer, which captures and processes the images on its own. The classifier using NT showed 83% accuracy while the ICM analyzer showed 73%. A combination of both methods yielded 84% accuracy.

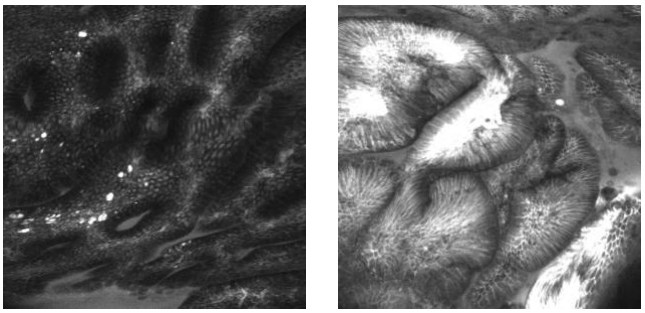
Münzenmayer et al [4], presents an algorithm for a content-based image retrieval (CBIR) framework. It uses color-texture analysis in order to determine images with and without lesions. In order to develop better queries, it uses a relevance algorithm to steer the search during the feedback loop. This paper measures its success using inter-rater reliability measurement of kappa = 0.71 [5] on a database of 390 images.

In Sabo et al [6], instead of analyzing endoscopic images the analysis is done on biopsies. From the given biopsies, computerized morphometry is used to extract attributes such as size, shape, texture, symmetry and architectural distribution of the epithelial nuclei.

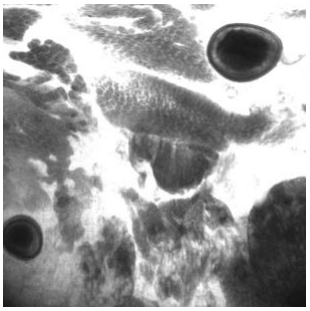
These attributes were used to classify the images as negative for dysplasia (ND), low-grade dysplasia (LD) and high-grade dysplasia (HD). Based on these attributes and their characteristics for each category, a two-layered feed forward back-propagation neural network was developed. It used a dataset of 152 self-collected images where 97 were used for training and 55 for evaluation. The neural network was able to differentiate ND and LD with an accuracy of 89% and differentiate LD and HD with an accuracy of 86%.

Qi et al. [7] used a CAD system for classification of dysplasia in Barrett´s esophagus.

Through endoscopic optical coherence tomography this CAD analyzes the image executing its four modules: region of interest segmentation, dysplasia-related image feature extraction, feature selection and site classification and validation. This software resulted in 84% of accuracy when classifying dysplastic vs. non-dysplastic Barret´s esophagus tissue.



GMP Image BAR Image



NPL Image

Figure 1 - Example CLE Images

Aubreville et al. [8] used Deep Learning technologies on Confocal Laser Endomicroscopy (CLE) images to automatically classify cancerous tissue in Laser endomicroscopy images of the oral cavity. The presented approach in CLE image recognition results in an area under the curve (AUC) of 0.96 and a mean accuracy of 88.3% (sensitivity 86.6%, specificity 90%).

Swager [9] developed a CAD system for the classification of Barrett´s cancer using Volumetric Laser Endomicroscopy (VLE). The evaluation was done using commonly used image analysis features (e.g. Local Binary Patterns - LBP, Histogram of Oriented Gradients - HOG and Gray-Level Co-occurrence Matrix - GLCM features) in combination with popular classification methods (e.g. Support Vector Machine - SVM, Random Forest and Neural Networks). Based on the clinical prediction model, in this study the author derived three features: (1) Layer histogram, (2) Large-scale gray-level co-occurrence matrix features and (3) Bin median of pixel averages, which were called clinically-inspired features. In this case, the proposed clinically-inspired features outperformed the state-of-the art alternatives and also human experts in distinguishing cancerous tissue from non-dysplastic tissue based on ex vivo VLE images.

Although the results presented by the analyzed papers were very promising, none of the techniques seems to be reliable enough to classify the exams and, in a real-world scenario, it would still be required that a biopsy be conducted on the patient to determine the presence of a disease.

IV. DATASET

In order to train, validate and attest the performance of the classifiers, a set of images of the diseases was required. This paper uses the image dataset from the **Analysis of Images to Detect Abnormalities in Endoscopy (AIDA-E)** challenge [19], which was published during the 2016 edition of **The IEEE International Symposium on Biomedical Imaging (ISBI)**. The proper permissions were given by the owners to use this dataset in this research. They also state that the dataset is open for use in research as long as the owners or the challenge itself is mentioned. The dataset consists of 262 images from CLE exams being 172 affected by BAR, 60 by NPL and 30 by GM. The images are 1024 pixels wide by 1024 pixels tall, with 8 bits for color.

In the deep learning approach used by Aubreville et al. [8] a dataset of more than 7 thousand images was used, reaching a mean accuracy of around 88%. Considering that the dataset available for this work is only composed of 262 images, it can be seen that deep learning approaches are very discouraged in the sense that they will very unlikely yield any significant results. Besides that, it can also be seen that the number of images for each disease is unbalanced, there being a lot more images for BAR than there are for NPL and GMP. This makes neural networks trained with these images biased towards BAR, which is not desirable.

To work around these issues, the images were organized into smaller groups. This was done by splitting the images from each category into sets of 30 images, resulting in 5 BAR sets, 2 NPL sets and one GMP set. By combining a set of each category, 10 distinct groups of 90 images each were generated. They were used as the training and validation data for each of the developed classifiers, with a leave-one-out strategy.

V. DEVELOPED METHODS

The objective of the proposed method is to classify the images in the dataset as being GMP, BAR or NPL.

In every experiment conducted with the classifiers, to extract attributes and analyze the images, the Scikit Image library [10] was used. The Scikit Learn [11] library provided the fundamental methods for designing, training and validating the classifiers.

The following sections describe a baseline classifier, followed by a classifier using GLCM attributes and one using LBP attributes.

A. Baseline Classifier

Before exploring attributes and techniques related to texture analysis, a baseline classifier was designed. This was done in order to have results to which the GLCM and LBP classifiers could be compared to.

By observing the average histogram of each category (generated by averaging the histograms of all images in the category), as seen in Figure 2, Figure 3 and Figure 4 it can be noted that they present characteristics in their distributions that can be used in order to classify them. Due to that, color intensity distribution was chosen as the source of the attributes in the baseline classifier.

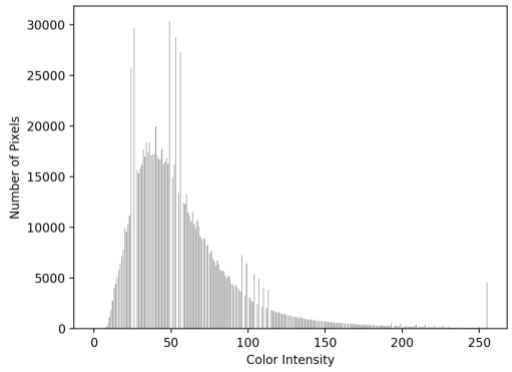


Figure 2 - Average Histogram for GMP

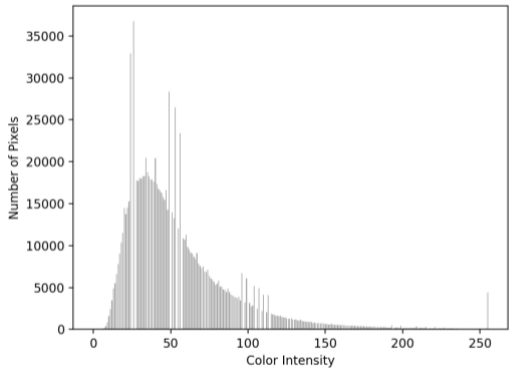


Figure 3 - Average Histogram for BAR

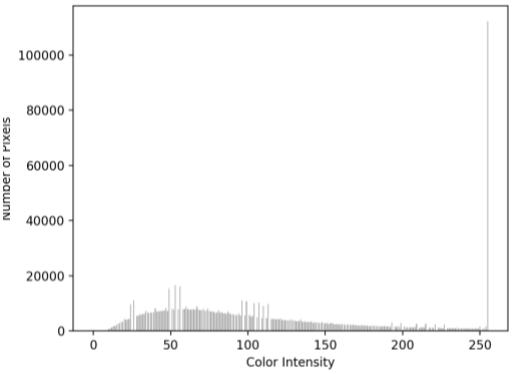


Figure 4 - Average Histogram for NPL

The baseline classifier was built using a neural network based on the Multi-Layer Perceptron (MLP) [12]. The network is composed of a single hidden layer of 25 neurons.

The inputs for this network are based on the color intensity of each pixel. The inputs are:

- **Mean** – Average of the grayscale intensities of each pixel in a given image;
- **Variance** – How the grayscale intensities of each pixel are distributed around the median;

Table 1 - Per Group Results for Baseline Classifier

Actual	GMP			Acc.	BAR			Acc.	NPL			Acc.
Predicted	**GMP**	**BAR**	**NPL**		**GMP**	**BAR**	**NPL**		**GMP**	**BAR**	**NPL**	
Group 1	13	14	3	**43%**	13	13	4	**43%**	1	3	26	**87%**
Group 2	17	12	1	**56%**	11	16	3	**53%**	1	3	26	**87%**
Group 3	22	6	2	**73%**	8	19	3	**63%**	4	3	23	**77%**
Group 4	19	10	1	**63%**	8	19	3	**63%**	4	6	20	**67%**
Group 5	16	11	3	**53%**	13	12	5	**40%**	5	4	21	**70%**
Group 6	19	8	3	**63%**	8	17	5	**57%**	2	5	23	**77%**
Group 7	13	13	4	**43%**	11	15	4	**50%**	4	1	25	**83%**
Group 8	16	11	3	**53%**	15	13	2	**43%**	2	1	27	**90%**
Group 9	21	8	1	**70%**	7	20	3	**67%**	4	2	24	**80%**
Group 10	15	11	4	**50%**	11	17	2	**57%**	3	2	25	**83%**
Total	**56.7% Mean Accuracy, Std. Dev. 10.38**				**53.6% Mean Accuracy, Std. Dev. 9.44**				**80.1% Mean Accuracy, Std. Dev. 7.4**			

- **Asymmetry** – Describes the concentration of values related to the median;
- **Kurtosis** – Describes the degree of flattening of the distribution function of the grayscale values in the image;
- **Pixels within the range 150:199** – Number of pixels with intensities between 150 and 199;
- **Pixels more intense than 200** – Number of pixels with intensities greater than (or equal to) 200.

The classifier training and validation was conducted using the groups of images generated from the dataset, described on section IV.A, and the **Leave One-Out Cross Validation** approach in which, for each group, the classifier is trained with all but one image, which is used to validate the classifier's performance afterwards. This process repeats once for every image in the group. The predictions given by the classifier for each image are then combined into the group's results. The classifier yielded the overall results shown in Table 2. Results for each group are shown in Table 2. The classifier shows a mean accuracy of 63.5%.

Table 2 - Overall results for the baseline classifier

	Pred. GMP	**Pred. BAR**	**Pred. NPL**
Actual GMP	171 (57%)	104 (35%)	25 (8%)
Actual BAR	105 (35%)	161 (54%)	34 (11%)
Actual NPL	30 (10%)	30 (10%)	240 (80%)

B. Classifier using Gray Level Co-Occurrence Matrix (GLCM)

The **Gray Level Co-Occurrence Matrix (GLCM) [13]** is a classic technique used to describe an image's texture, i. e., the relationship between neighboring pixels, instead of just their intensities as an isolated information.

A GLCM is calculated by tabulating the occurrences of a pair of pixel intensities in a given distance and direction. As seen in Figure 5, the number of times two pixels are positioned as described by *a* and *b* is accounted for in *a*'s intensity's line and *b*'s intensity's column of the matrix.

The matrices in the figure, starting in the upper-right corner, clockwise, are 0° with distance 1, 45° with distance 1 and 270° with distance 1. After all pairs have been accounted for, all of the numbers are then divided by the total number of pairs found in the image. By doing so it becomes possible to read the value as the probability of a pair showing up in the original image.

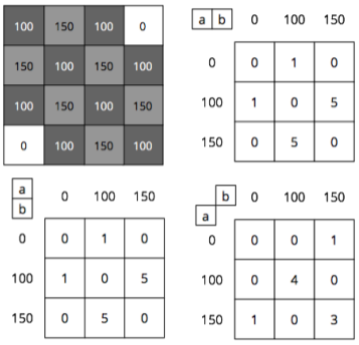


Figure 5 – Example of GLCMs of an image

The extraction of attributes from GLCM was first introduced by Haralick et al. in [16][8], however, Baraldi and Parmiggiani [17] show that only some of them are actually relevant when describing textures, which are:

- Angular Second Moment
- Entropy
- Contrast
- Variance
- Correlation
- Homogeneity

For each image, four matrices were calculated with a pixel distance of 1 and angles of 0°, 45°, 90° and 135°. The six attributes aforementioned were extracted from these matrices and used as inputs for the classifier totaling, along with the six from the baseline classifier, 30 distinct inputs. The training and validation were conducted using the same approach and image groups as the baseline classifier. The classifier yielded the overall results shown in Table 4. Results for each group are shown in Table 3. The classifier shows a mean accuracy of 67.5%.

Table 3 - Group Results for GLCM Classifier

Actual	GMP			Acc.	BAR			Acc.	NPL			Acc.
Predicted	GMP	BAR	NPL		GMP	BAR	NPL		GMP	BAR	NPL	
Group 1	18	10	2	**60%**	7	17	6	**57%**	3	4	23	**77%**
Group 2	16	11	3	**53%**	13	16	1	**53%**	1	3	26	**87%**
Group 3	20	6	4	**67%**	8	20	2	**67%**	3	4	23	**77%**
Group 4	19	5	6	**63%**	5	23	2	**77%**	3	1	26	**87%**
Group 5	17	11	2	**57%**	9	14	7	**47%**	3	7	20	**67%**
Group 6	17	12	1	**57%**	9	18	3	**60%**	1	3	26	**87%**
Group 7	18	9	3	**60%**	9	17	4	**57%**	3	3	24	**80%**
Group 8	17	11	2	**57%**	7	22	1	**73%**	4	1	25	**83%**
Group 9	16	12	2	**53%**	7	19	4	**63%**	3	1	26	**87%**
Group 10	15	9	6	**50%**	6	22	2	**73%**	1	0	29	**97%**
Total	**57.7% Mean Accuracy, Std. Dev. 5**				**62.7% Mean Accuracy, Std. Dev. 9.7**				**82.8% Mean Accuracy, Std. Dev. 8.1**			

Table 4 - Overall Results for GLCM Classifier

	Pred. GMP	Pred. BAR	Pred. NPL
Actual GMP	173 (58%)	96 (32%)	31 (10%)
Actual BAR	80 (27%)	188 (63%)	32 (11%)
Actual NPL	25 (8%)	28 (9%)	247 (82%)

C. Classifier Using Local Binary Patterns (LBP)

The **Local Binary Patterns (LBP) [16]** like the GLCM, is another classic technique used to describe an image's texture. The LBP matrix is calculated by, first, defining a distance (*d*) and an amount of points (*n*). Then, for every pixel in the image, *n* points are taken from the neighboring pixels at distance *d*. If the intensity of that pixel is greater than the current one's, a '1' is written. Otherwise, a '0' is written.

After all the pixels have been computed, a string of '1's and '0's will have been generated. That string is then used as a binary number and converted into a decimal number, which is placed in the according position in the LBP matrix. The generated matrix will have the same size as the original image.

For example, considering the case with the distance equal to 1 pixel, taking the upper-right bit as the first one and using a clockwise direction, the generated binary code would be 11110000, which is represented by the decimal number 240, after a standard binary conversion.

Out of the generated matrix it is possible to extract statistical attributes such as: Mean, Variance, Asymmetry and Kurtosis; which were explored in the baseline classifier.

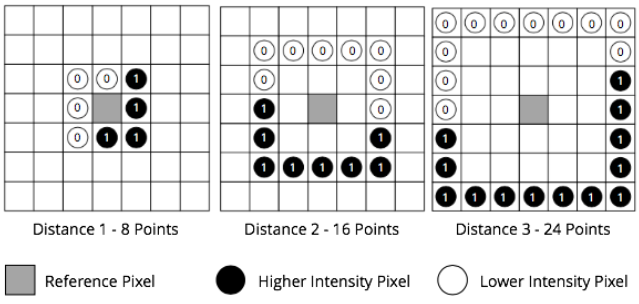


Figure 6 – LBP Matrices

This classifier uses 3 LBP matrices to generate the inputs: one with distance 1 and 8 points, one with distance 2 and 16 points and one with distance 3 and 24 points, as seen in Figure 6. The four statistical attributes are calculated from these matrices, totaling, along with the baseline attributes, 18 attributes which are used as input.

The training and validation were, again, conducted using the same approach and image groups as the baseline classifier. The classifier yielded the overall results shown in Table 5. Results for each group are shown in Table 6. The classifier shows a mean accuracy of 67.7%.

Table 5 - Overall Results for LBP Classifier

	Pred. GMP	Pred. BAR	Pred. NPL
Actual GMP	187 (62%)	89 (30%)	24 (8%)
Actual BAR	86 (29%)	176 (59%)	38 (12%)
Actual NPL	12 (4%)	41 (14%)	247 (82%)

VI. CONCLUSIONS

In this paper we explored the impact of adding attributes related to classic texture description techniques to a classifier for separating images of GMP, BAR and NPL. This analysis was done using a small and unbalanced dataset, which discouraged deep learning approaches. By creating smaller balanced groups of images, we were able to train and validate the classifier without biasing it towards any of the outcomes.

It could be seen in the baseline classifier, which relied only on color intensity attributes, that the results were only reasonable at 63.5% mean accuracy and that the standard deviation for the group accuracies was considerably high at around 9 points. In the GLCM classifier, an increase of 4% in mean accuracy could be observed and, in the LBP classifier, 4.2%.

Although the accuracies for both the GLCM and LBP classifiers were pretty similar, it can be noted that the standard deviation for group accuracies in the GLCM classifier still roamed around 7,6 points, while for the LBP classifier the standard deviation lowered to around 4,7 points. This shows that both techniques have a similar impact in accuracy, but the use of LBP resulted in more consistent results for the groups. It

Table 6 - Group Results for LBP Classifier

Actual	GMP			Acc.	BAR			Acc.	NPL			Acc.
Predicted	**GMP**	**BAR**	**NPL**		**GMP**	**BAR**	**NPL**		**GMP**	**BAR**	**NPL**	
Group 1	17	11	2	**57%**	8	18	4	**60%**	1	4	25	**83%**
Group 2	21	9	0	**70%**	10	19	1	**63%**	1	4	25	**83%**
Group 3	17	9	4	**57%**	7	16	7	**53%**	1	7	22	**73%**
Group 4	19	10	1	**63%**	8	19	3	**63%**	2	4	24	**80%**
Group 5	19	8	3	**63%**	7	17	6	**57%**	0	7	23	**77%**
Group 6	19	9	2	**63%**	9	19	2	**63%**	0	4	26	**87%**
Group 7	20	9	1	**67%**	9	15	6	**50%**	3	3	24	**80%**
Group 8	19	7	4	**63%**	11	16	3	**53%**	1	3	26	**87%**
Group 9	18	8	4	**60%**	10	17	3	**57%**	1	3	26	**87%**
Group 10	18	9	3	**60%**	7	20	3	**67%**	2	2	26	**87%**
Total	**62.3% Mean Accuracy, Std. Dev. 4.0**				**58.6% Mean Accuracy, Std. Dev. 5.5**				**82.4% Mean Accuracy, Std. Dev. 4.8**			

can also be seen that the LBP classifier reached the same mean accuracy as the GLCM classifier while using less attributes as input.

When compared to the aforementioned researches on similar subjects, some interesting differences come up. Looking at [2] we can see that the number of inputs for the classifier is relative to the image size, since it computes 2 attributes for every 50x50 pixel block. That means that, for images with different sizes, the classifier would have to be remodeled. In the approaches presented in this paper, all the classifiers require a specific amount of inputs, for any given image size, requiring only that the model be retrained, not remodeled, for a different image size.

The deep-learning approach taken in [8] requires a much larger dataset of images in order to be trained. That dataset was composed of more than 7 thousand images, whereas the classifiers presented here were trained with 90. Considering that a mean accuracy of 82.8% was achieved for NPL images, there is a reason to believe that a classifier can reach similar results to the ones in [8] even with smaller datasets.

Overall, the use of attributes taken from texture description techniques can be seen as beneficial for improving the accuracy of image classifiers and, the use of LBP in particular, for improving the consistency of the results. Even with a small dataset and the issue of the unbalance between categories, these techniques still showed significant improvement.

VII. REFERENCES

[1] Salomao, M. A., Lam-Himlin, D., & Pai, R. K. (2017). **Substantial Interobserver Agreement in the Diagnosis of Dysplasia in Barrett Esophagus Upon Review of a Patient's Entire Set of Biopsies**. The American Journal of Surgical Pathology, 1.

[2] van der Sommen, F., Zinger, S., Curvers, W. L., Bisschops, R., Pech, O., Weusten B. L. A. M., Bergman, J. J. G. H. M., de With, P. H. N., Schoon, E. J. (2016). **Computer-aided detection of early neoplastic lesions in Barrett's esophagus**. Endoscopy, 48, 617-624.

[3] Dunn, J. M., Hveem, T., Pretorius, M., Oukrif, D., Nielsen, B., Albregtsen, F., . . . Danielsen, H. E. (2011). **Comparison of nuclear texture analysis and image cytometric DNA analysis for the assessment of dysplasia in Barretts oesophagus**. British Journal of Cancer, 105(8), 1218-1223.

[4] Münzenmayer, C., Kage, A., Wittenberg, T., & Mühldorfer, S. (2009). **Computer-assisted Diagnosis for Precancerous Lesions in the Esophagus**. Methods of Information in Medicine, 48(4), 324-330.

[5] Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G. J. Miller, D. H. (1996). **Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques**. Magnetic resonance imaging, 14 (5), 495–505.

[6] Sabo, E., Beck, A. H., Montgomery, E. A., Bhattacharya, B., Meitner, P., Wang, J. Y., & Resnick, M. B. (2006). **Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in Barretts esophagus**. Laboratory Investigation, 86(12), 1261-1271.

[7] Qi, X., Sivak, J. M., Wilson, D. L., & Rollins, A. M. (2004). **Computer-aided diagnosis of dysplasia in Barrett"s esophagus using endoscopic optical coherence tomography**. Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine VIII.

[8] Aubreville, M., Knipfer, C., Oetter, N., Jaremenko, C., Rodner, E., Denzler, J., … Maier, A. (2017). **Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning**. Scientific Reports, 7, 11979.

[9] Swager A (2017). **Volumetric laser endomicroscopy for the detection of early Barrett's neoplasia.** University of Amsterdam, 236 Pages.

[10] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu and the scikit-image contributors. **scikit-image: Image processing in Python**. PeerJ 2:e453 (2014)

[11] Pedregosa et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. V12, 2825-2830.

[12] J. Han, M. Kamber (2011). **Data mining: concepts and techniques**. 3rd ed Burlington, MA:Elsevier, 703p.

[13] Campbell, D. L., Kang, H., & Shokouhi, S. (2017). **Application of Haralick texture features in brain [^{18}F]-florbetapir positron emission tomography without reference region normalization**. Clinical Interventions in Aging, Volume 12, 2077-2086.

[14] Robert M. Haralick, K. Shanmugam, Its'Hak Dinstein (1973). **Textural Features for Image Classification.** IEEE Transactions on Systems, Man, and Cybernetics. Vol. SMC-3, No. 6. (November 1973), pp. 610-621

[15] Baraldi, A., & Parmiggiani, F. (1995). **An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters**. IEEE Transactions on Geoscience and Remote Sensing, 33(2), 293-304.

[16] T. Ojala, M. Pietikinen, T. Maenpaa (2002). **Multiresolution gray-scale and rotation invariant texture classification with local binary patterns**, IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971-987.

[19] **Analysis of Images to Detect Abnormalities in Endoscopy (AIDA-E) challenge:** ttps://isbi-aida.grand-challenge.org

[20] Muhammad W. Shahid, Michael B. Wallace,**Endoscopic Imaging for the Detection of Esophageal Dysplasia and Carcinoma**. Gastrointestinal Endoscopy Clinics of North America,Volume 20, Issue 1,2010,Pages 11-24