Classification of lymphomas images with polynomial strategy: An application with Ridge regularization

Danilo C. Pereira* §, Leonardo C. Longo†, Thaína A. A. Tosta‡, Alessandro S. Martins§, Adriano B. Silva*, Paulo R. de Faria¶, Leandro A. Neves†, Marcelo Z. do Nascimento*
*Faculty of Computer Science, Federal University of Uberlandia, Minas Gerais, Brazil
†Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP)
‡Science and Technology Institute, Federal University of São Paulo (UNIFESP)
§Federal Institute of Triângulo Mineiro (IFTM)
¶Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia (UFU)
E-mails: *danilo.pereira@ufu.br*, *leonardohcl@hotmail.com*, *tosta.thaina@unifesp.br*, *alessandro@iftm.edu.br*, *adriano.barbosa@ufu.br*, *paulo.faria@ufu.br*, *leandro.neves@unesp.br*, *marcelo.nascimento@ufu.br*

***Abstract*—Histological image analysis through systems to aid diagnosis plays an important role in medicine with supplementary reading for the specialist's diagnosis. This work proposes a method based on the association of extracted features by fractal techniques, regularization and polynomial classifier. The feature vectors were classified by applying the cross-validation technique with 10 folds. The evaluation of the results occurred through metrics such as accuracy (ACC) and imbalance accuracy metric (IAM). The proposed approach achieved significant results for all metrics with non-Hodgkin lymphoma lesion sets. The proposed approach provided values around 0.97 of IAM and 99% of ACC for investigated groups. These results are considered relevant to studies in the literature and the association of Hermite polynomial and regularization can contribute to the detection of the lesions by supporting specialists in clinical practices.**

***Keywords*—CAD, Histological Image, Polynomial Classifier, Regularization.**

I. Introduction

Lymphoma is a type of cancer that affects the lymphatic system and can be subclassified into 38 types according to their morphological, immunophenotypic, genetic, and clinical characteristics [1]. One of the types with the highest evidence among diagnosed clinical cases, around 85%, is non-Hodgkin lymphomas (NHL). This type can be identified as chronic lymphocytic leukemia (CLL), follicular lymphoma (FL) and mantle cell lymphoma (MCL) [2].

Identification and diagnosis of NHL by the pathologist is still a challenging and complex task. Factors such as workload and specialist experience level can influence this process. To assist the specialists, computer-aided diagnosis (CAD) systems have been applied for supplementary reading during tissue analysis. The main steps of these systems include the acquisition of the digitized image, preprocessing, segmentation of regions of interest (ROIs), feature extraction, and classification. According to the authors [3], the development of feature extraction and classification algorithm capable of identifying the ROIs with information on this type of cancer is fundamental in the diagnosis [4].

The feature extraction should bring information that can provide differentiation of the labels in the classification step. Among the approaches, an important technique for description is the quantification of the texture of an image. Texture provides measurements of properties such as smoothness, roughness, and regularity of the image. Techniques based on fractal dimension and lacunarity allow quantifying self-similarity properties present in the images, which can not be defined by Euclidean geometry. These properties are fundamental to the processes employed in the classification of different types of NHL [5].

Classification is a machine learning step that employs the feature set as the basis for assigning class labels to patterns obtained from input images [6]. Although various methods contribute to solving biomedical classification problems, the choice of the appropriate technique for this task depends on the properties of the classification algorithm [7]. The polynomial classifier is an algorithm that has shown promising results, especially when dealing with nonlinearly separable data. This algorithm is a parameterized method that exponentially expands its polynomial base according to the number of elements in the data vector and the degree of function. So, this strategy can increase the processing time and this fact represents an important challenge in the classification stage [8]. So, a regularization method can provide stability to the classification and that makes it less sensitive to data training processes. Regularization may also lead to a reduction in the generalization error as well as also help in feature selection to find out the most relevant features for the classification process [9]. Several studies have explored solutions as described in these works [10], [11]. However, a method that associates regularization with the polynomial classifier has not yet been investigated, especially when data of NHL histological images are considered.

This study presents an approach for NHL lesion classification based on descriptors obtained by fractal geometry and polynomial classifier. The feature extraction step employed the fractal geometry approach for vector generation considering LAB and RGB color models. The Ridge regularizer was employed to equalize the set of attributes by adding a constant that promotes the resizing of the data. Finally, these vectors were labeled using the polynomial classifier with the Hermite base function. The performance of the proposed method was evaluated by the imbalance relation between accuracy metric (IAM) [12] and accuracy (ACC). The polynomial classifier was also compared with machine learning (ML) algorithms. The contributions of this work are i) Study of the polynomial algorithm based on the Hermite base function with a new parameterization for NHL classification; ii) Investigation of Ridge regularization technique on descriptors obtained from fractal geometry and the performance on polynomial classification with color models; iii) Performance comparison of the Hermite polynomial algorithm with ML methods using the IAM metric.

II. Methodology

The proposed work performs the classification of lesions through algorithms developed in the Python programming language and WEKA software, as presented in Fig. 1. The experiments were evaluated on a computer with an Intel Core i5 with 16 GB RAM.

A. Database

The proposed methods were evaluated with data from the National Cancer Institute and National Institute on Aging, in the United States [13]. The dataset was constructed from 30 histological slides of lymph nodes stained with H&E, scanned using the RGB color model with 24-bit quantization and resolution of 1388 $\times$ 1040 pixels and stored in TIFF format. The regions of each slide were digitally obtained using a light microscope (Zeiss Axioscope) with a 20$\times$ objective lens and an AxioCam MR5 coupled and recorded without compression. A total of 374 images were obtained, where 113 of CLL, 139 FL and 122 MCL.

B. Feature Extraction

The first stage consists of feature extraction, a process in which transformations are applied to image pixels to generate numerical values that are significant for pattern discrimination [14]. The used techniques are fractal dimension and lacunarity as demonstrated in [11] which were applied on RGB and CIE-LAB color model images. The LAB color model was chosen due to its compatibility with human vision, which makes it possible to represent all colors visible to the human eye. The representation used to define the RGB hyperspace was applied to evaluate the CIE-LAB model. One of the main advantages of fractal geometry-based approaches is the possibility of representing structures captured by the human visual system in a format that computer systems can quantify in more detail than Euclidean geometry [5]. In this work, the fractal geometry metrics were calculated using the gliding box method with the size of the box between $L = 3$ and $L = 45$.

According to the authors in [5], the investigation of the fractal dimension descriptor with different color models is a relevant combination because there are few approaches to the calculation of fractal measurements for color images. Based on the information described in the study [15], the use of separate color channels of different models can provide more relevant information in discriminating these types of lesions.

The lacunarity was extracted by quantification of the distribution and organization of the pixels as demonstrated in [16]. Therefore the feature vectors stored the features: fractal dimension metrics (FD) and five lacunarity values (Lac) (area under curve - ARC, skewness - SKW, area ratio - AR, maximum point - MP and scale of the maximum point - SMP), represented by Lac (1) to Lac (5). These measurements were obtained from each channel separated from the color models, totaling 18 features in the data vector.

The ARC metric was obtained through numerical integration using the trapezoidal method. The SKW consisted of an asymmetry indication compared to the average value. The negative values of skewness indicated that the sample was concentrated to the left of the average value. The positive values of SKW indicated that the sample was concentrated to the right of the average value. Considering a case of a perfectly symmetrical sample the skewness obtained is 0. The AR considered the ratio between the right side and the left side areas under the function curve. The MP provided the value of the maximum point of each function and the SMP is the scale of the maximum point.

C. Regularization

Regularization is a technique developed to standardize the attributes by adding terms that promote resizing and make the problem more stable. This approach allows solving the problem of the number of features superior to the number of classes since some input variables can be set to zero, providing the construction of a more efficient predictive model [17]. In this work, we applied the regularization method for feature selection and evaluation of the classification algorithms. The regularization stage was used in each training fold to obtain the most relevant features for the classification process. This allowed the reduction of the number of attributes required in the classification step [18]. These selected features were applied to all investigated classifiers.

An objective function *F* is smoothed by the norm penalty parameter $\omega(\theta)$, according to Eq. 1.

$$R(\theta, X, y) = F(\theta, X, y) + \beta\omega(\theta), \tag{1}$$

where X is the input matrix, y is the target labels, θ denotes the trainable parameters, $\beta \in [0, \infty)$ is a hyperparameter that weights the relative contribution of the norm penalty term ω, which is relative to the standard objective function F. When the β value is equal to 0 there is no penalty. However, as this parameter increases, the greater the adjustment performed.

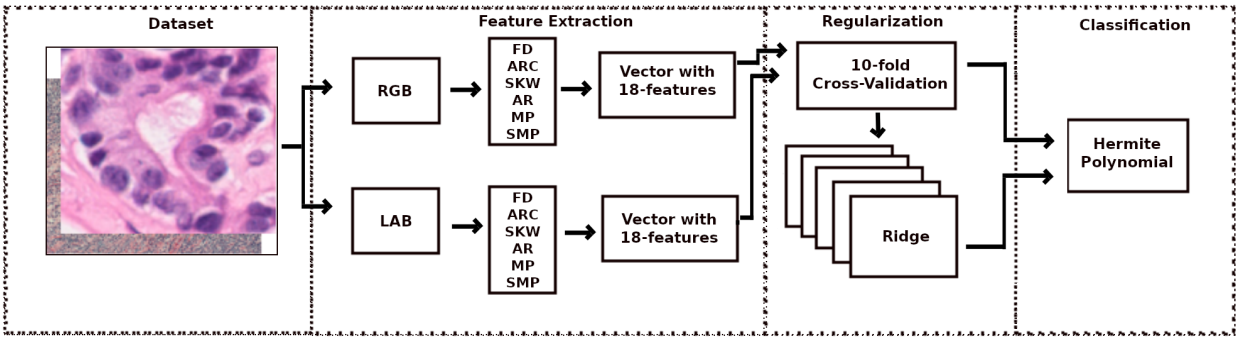

Fig. 1. Block diagram of the main stages of the proposed approach.

The Ridge regularization, represented by (Eq. 2), is obtained by adding to the cost function the sum of the squares of the weights β. The term $\lambda(0 < \lambda < \infty)$ is a hyperparameter that controls the application of the regularization method on each attribute to penalize them with different values

$$\underbrace{F(\theta, X, y)}_{\text{Original Cost Function}} + \underbrace{\lambda \sum_{j=1}^{p} \beta_j^2}_{\text{Ridge Regularization}} . \quad (2)$$

The value of λ that generates the highest classification rate can be calculated iteratively from a cross-validation process. For this approach, the range used was 10^{-5} to 10^{-3} with $2.5{\times}10^4$ iterations. These values were obtained empirically over the feature vector.

D. Classification

The polynomial classifier is a supervised method capable of nonlinearly expanding an input feature vector to a higher dimension. This strategy allows obtaining linear approximations in this space that can label the input data to the desired output [19].

The Hermite polynomials (HP) can generate a complete orthogonal basis of the Hilbert space that satisfy the orthogonality and completeness conditions of the family of elements of that space [20]. The orthogonality condition shown on Eq. 3, implies that any inner product between a pair of orthogonal polynomials, $P_m(x)_{m=0}^{\infty}$ and $P_n(x)_{n=0}^{\infty}$, of different degrees is equal to zero when the used base function $w(x)$ and $[a, b]$ range are the same [21].

$$< P_m(x), P_n(x) >= \int_a^b P_m(x) P_n(x) w(x) dx = 0, \forall m \neq n. \quad (3)$$

According to Thangavelu [22], HP is orthogonal in the range of $(-\infty, \infty)$, for the base function, and allows gains in the approximation of functions. Mathematically, HP can be defined as Eq. 4:

$$H_n(x) = (-1)^n e^{-x^2/2} \frac{d^n}{dx^n} [e^{-x^2/2}]. \quad (4)$$

An efficient way from a computational point of view to obtain this polynomial is by using the recurrence relation. In this case, only the first two terms are needed so that the others are iteratively calculated. The recurrence relation for the HP is expressed by Eq. 5 [6].

$$H_{n+1}(x) - xH_n(x) + nH_{n-1}(x) = 0, \quad n = 0, 1, 2, \ldots \quad (5)$$

Finally, the classifier can be defined according to the Eq. 6:

$$g(\mathbf{x}) = \mathbf{a}^T H_n(\mathbf{x}), \quad (6)$$

wherever **a**, is the coefficient vector of the polynomial base function, $H_n(\mathbf{x})$ is the *Hermite* base function and n corresponds to the order or degree. The training on the polynomial classifier enlarges the data vector through polynomial expansion to increase the separation of the classes in the generated new vector. This process generates a polynomial function that allows the separation of the data into two classes. In the testing step, the polynomial function generated earlier is applied to define the class to which each sample belongs [23]. In this work, the order of the polynomial base was empirically defined and the best results were obtained with the $3th$ order and five features for class separation.

E. Evaluation of Methods

The performance of the HP algorithm was compared with three different classifiers that are based on the main supervised ML approaches function-based, ensemble learning, and tree-based. The chosen algorithms were: logistic (LGT), a model based on the integration of tree induction algorithm and additive logistic regression; multilayer perceptron (MLP), an approach that consists of a system of simple interconnected neurons, or nodes, representing a nonlinear mapping between an input vector and an output vector; and random forest (RF), um strategy based on a tree ensemble method, wherein the bootstrapping for each tree yields subsets of observations which are not included in the tree growing process. These algorithms were evaluated with the cross-validation method *k*=10, wherever 90% of the dataset was used for training and 10% for testing the model. This study evaluated the groups of NHL lesions (FL-CLL, FL-MCL, and CLL-MCL). The ML

algorithms were employed from the Weka software and the default parameters were adopted.

The classification algorithms were evaluated using the IAM and ACC. The ACC metric is among the most common metrics for analysis of results in the context of image classification because the calculation is simple and the interpretation of the results is easy. This metric corresponds to a value between 0 and 1. If the classifier has correctly labeled all samples, the accuracy will have a value equal to 100%. However, in problems where classes are unbalanced, accuracy may not be a reliable measure. Due to some of the existing metrics recall of results based on the feature of the data for unbalanced labels, in this work the IAM defined as Eq. 7 was also considered in the experiments.

$$IAM = 1/k \sum_{i=1}^{k} \frac{c_{ii} - max(\sum_{j \neq i}^{k} c_{ij}, \sum_{j \neq i}^{k} c_{ji})}{max(c_{.i}, c_{i.})}, \quad (7)$$

In computing the IAM, c_{ij} is the confusion matrix generated by the classifier. The max value of total off-diagonal items ($\sum_{j \neq i}^{k} c_{ij}$ or $\sum_{j \neq i}^{k} c_{ji}$) are subtracted from the diagonal values (c_{ii}), divided by the max sum in the corresponding row or column ($max(c_{.i}, c_{i.})$), and finally averaged ($/k$) to obtain the expectation.

III. Results and Discussion

Table I shows the results obtained with the IAM and ACC metrics for each group (FL-CLL, FL-MCL, and CLL-MCL) with the color channels of LAB and RGB models without using the regularization. As there is a group of studies in the literature that investigates NHL lesions in binary classification approaches, in this study the investigations followed the same strategy [24], [25]. According to the results in Table I, we can observe that the HP classifier was superior for both color models investigated by the ACC metric. The IAM metric values (see Table 1) show that the standard deviation values are also smaller with the HP algorithm. The IAM metric shows the number of instances that the classifier correctly predicted. We observed that the ML algorithms had a value closer to zero when compared with the HP classifier. This metric has a range of representation between -1 to 1 and when this value is closer to zero this data represents that number of correct and incorrect classified instances is close. Results with the MLP algorithm provided 84.2% of ACC to the CLL-MCL group, but when the IAM metric is analyzed we can see that the number of instances correct and incorrect was affected (0.40). Furthermore, we observed that the standard deviation values are higher for the ML algorithms in all groups. This is due to the great variation of the results obtained in each $fold$, especially with results obtained by the IAM metric. For instance, in the FL-CLL group, the standard deviation value was 0.04 for the HP algorithm and 0.22 for the LGT technique. According to the authors in [12], the IAM metric shows the behavior of the classifier is expected to not label an instance of data in incorrect classes.

Furthermore, in Table I, we observed that the IAM average value was 0.96 for LAB channels with the HP algorithm. For other algorithms, the average value were 0.42, 0.54, and 0.49, respectively, for the LGT, MLP, and RF classifiers. For the RGB channels, the average values were 0.96, 0.37, 0.50, and 0.44 with the HP, LGT, MLP, and RF algorithms, respectively. This shows the difference between the use of this model for the other classifiers. It was observed that the HP algorithm provided higher values than the other ML algorithms with the use of the IAM metric. The data in Table I indicate that the average value of ACC for the HP classifier was 98.70%, 75.30% for the LGT, 82.00% with the MLP, and 79.00% in RF. These values represent a difference of 24%, 17%, and 20%, among HP, LGT, and MLP classifiers, respectively.

Table II presents the results based on the study of the regularization method. On Table II, we observed that there is a decrease in values with the ACC and IAM metrics employed for evaluation of the HP classifier with the modified vectors. However, the HP method still retains the best results in comparison with other algorithms. The average value of IAM in the LAB channels for the HP, LGT, MLP and RF classifiers were 0.93, 0.38, 0.53, and 0.52, respectively. Using the RGB channel features, this metric presents the average values of 0.92, 0.35, 0.39, and 0.46 with the HP, LGT, MLP, and RF algorithms, respectively. The accuracy was around 97.00% for the HP classifier, 74.00% with LGT, 81.00% for MLP, and 80.00% with RF. These values represent an average difference amongst the algorithms compared and HP of 24%, 17%, and 16%, successively. In these experiments, only the RF classifier obtained improvement results for all groups in the RGB color model. However, these values were lower than those obtained with the HP algorithm.

Furthermore, we evaluated the frequency of the selected features by the Ridge algorithm. Figs. 2 and 3 show the frequency histogram of each selected feature with the LAB and RGB color models. The features identified from 1 to 18 in Fig. 2 correspond to FD, ARC, SKW, AR, MP and SMP for the L, A and B channels, respectively. Similarly, in Fig. 3, the attributes are arranged in the same order considering the R, G, and B channels. We observed that the DF measure extracted from channel L (see Fig. 2) appeared in all folds of the FL-CLL group. On the other hand, we observed that the DF value was not selected from the A channel. The most selected features in the folds were MP of the L channel, AR and MP for channel A and AR for channel B. These data were represented by the numbers 5, 10, 11, and 16 on the X axis of the graph. Thus, these four features had the greatest influence on the classification. On the other hand, in Fig. 3, the features that had the highest incidence were SKW, AR, MP, and SMP of the R and B channels, i.e., eight information that most contribute to the classification. This information is represented by the numbers 3, 4, 5, 6, 15, 16, 17, and 18 on the X axis of the graph. In some groups such as FL-CLL, some features were not selected in the folds (numbers 8 and 14 in Fig. 3). The performance of the HP algorithm remains promising even with the lower value when using the regularizer with the IAM metric. Therefore, this strategy can reduce the complexity of computing the algorithms. The best

TABLE I
RESULTS OBTAINED WITH CLASSIFICATION ALGORITHMS FOR THE DATASET WITH LAB E RGB WITHOUT REGULARIZATION

		FL-CLL		FL-MCL		CLL-MCL		Average	
Classifier	Colormap	IAM	ACC	IAM	ACC	IAM	ACC	IAM	ACC
HP	LAB	0.98±0.04	99.6±0.04	0.95±0.07	98.4±0.02	0.96±0.07	99.2±0.01	**0.96±0.06**	**99.1±0.02**
	RGB	0.98±0.04	99.6±0.01	0.97±0.06	99.2±0.01	0.93±0.08	97.9±0.03	**0.96±0.06**	**98.9±0.02**
LGT	LAB	0.45±0.22	78.1±0.12	0.41±0.18	75.8±0.07	0.41±0.11	76.7±0.07	0.42±0.17	76.9±0.09
	RGB	0.45±0.27	77.3±0.10	0.42±0.18	76.5±0.12	0.24±0.07	67.5±0.11	0.37±0.17	73.8±0.11
MLP	LAB	0.67±0.18	86.2±0.07	0.57±0.13	83.6±0.07	0.40±0.24	84.2±0.09	0.55±0.18	84.7±0.08
	RGB	0.60±0.19	85.4±0.06	0.54±0.11	80.8±0.06	0.36±0.15	72.7±0.07	0.50±0.15	79.7±0.06
RF	LAB	0.50±0.20	81.7±0.10	0.50±0.16	79.7±0.07	0.48±0.13	82.5±0.06	0.49±0.16	81.3±0.08
	RGB	0.48±0.16	81.0±0.05	0.45±0.14	75.8±0.07	0.40±0.14	74.9±0.06	0.44±0.14	77.2±0.06

TABLE II
RESULTS OBTAINED WITH CLASSIFICATION ALGORITHMS FOR THE DATASET LAB AND RGB WITH REGULARIZATION

		FL-CLL		FL-MCL		CLL-MCL		Average	
Classifier	Colormap	IAM	ACC	IAM	ACC	IAM	ACC	IAM	ACC
HP	LAB	0.97±0.06	99.2±0.01	0.91±0.07	97.2±0.02	0.91±0.08	98.0±0.02	**0.93±0.07**	**98.1±0.02**
	RGB	0.96±0.06	97.5±0.02	0.91±0.01	97.6±0.01	0.90±0.11	97.1±0.01	**0.92±0.06**	**97.4±0.02**
LGT	LAB	0.33±0.26	73.6±0.11	0.41±0.20	75.8±0.09	0.41±0.16	74.9±0.07	0.38±0.20	74.8±0.09
	RGB	0.44±0.26	78.2±0.10	0.39±0.13	74.2±0.07	0.24±0.23	67.5±0.11	0.35±0.20	73.3±0.09
MLP	LAB	0.69±0.13	87.9±0.06	0.57±0.10	83.2±0.05	0.34±0.19	75.8±0.08	0.53±0.14	82.3±0.06
	RGB	0.58±0.11	85.0±0.04	0.33±0.13	81.3±0.05	0.27±0.17	71.4±0.08	0.39±0.29	79.2±0.06
RF	LAB	0.53±0.28	80.9±0.13	0.53±0.12	80.5±0.04	0.51±0.20	80.1±0.08	0.52±0.20	80.5±0.08
	RGB	0.55±0.17	83.4±0.06	0.50±0.16	80.8±0.08	0.34±0.16	75.8±0.08	0.46±0.16	80.0±0.07

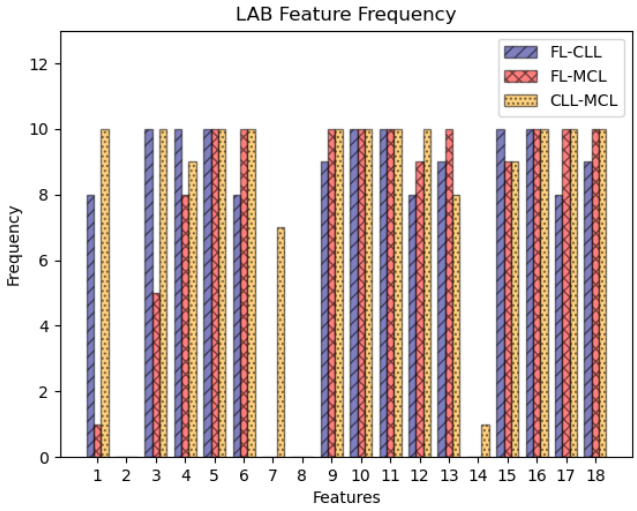

Fig. 2. Frequency of regularized features in the LAB color model.

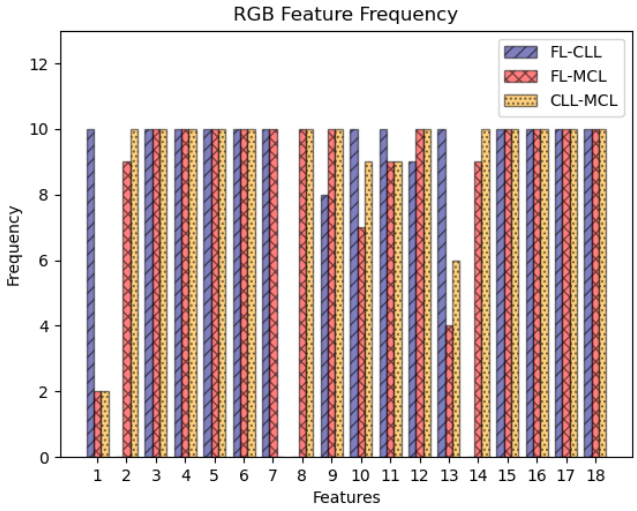

Fig. 3. Frequency of regularized features in the RGB color model.

performance with the reduction of the feature number was observed by the HP classifier in the FL-MCL group and with the LAB color model. To obtain this result, the proposed approach used in part of the experiments the following features: MP and SMP of the L channel; SWK, AR and MP from channel A and FD, AR, MP, and SMP from channel B. An analysis of computational time was also carried out on the HP classifier without applying the regularization. In this experiment, the HP algorithm provided an average time of 24,840 seconds. With the regularization method, the average time was 1,540 seconds which corresponds to a reduction of 16 times.

Table III shows an indirect comparison with recently published works that proposed different techniques for NHL image classification. The results show that the approach with the relevant accuracy value for NHL classification also used several different features. Even so, the HP algorithm together with the fractal descriptors provided promising results when analyzing other works in the literature.

IV. CONCLUSION

The great impact generated by the work was to investigate the association of fractal features, regularization, and HP for the construction of prediction and classification models of

TABLE III
AN INDIRECT COMPARISON OF THE ACCURACY FROM DIFFERENT APPROACHES

Reference	Approach	ACC(%)
Proposed	Fractal Dimension, Lacunarity Polinomial Classifier and Regularization	98.13
[24]	Morphology, entropy, GLCM, other descriptors, ELM and SVM	97.96
[8]	Fractal Dimension, Lacunarity and Polynomial Classifier	97.60
[26]	Percolation theory and Logistic Classifier	96.40
[25]	Histogram, edge histogram, LBP, curvelets, colour, wavelet and SVM	95.50

NHL tissues concerning digital histology. The results showed that the proposed method provides a good discriminative performance in RGB and LAB color models with the IAM and ACC metrics above 0.90 and 97%, which are relevant results when indirectly compared to the literature. Furthermore, this strategy shows that a regularization is a relevant tool for computational time reduction spent by HP for model training.

The limitation of this work is related to the restricted group of samples which is believed to have influenced the results of the experiments. In the future studies will be applied to principal component analysis to verify the correlation of the selected features by the regularization. In addition, it aims to increase the number of NHL samples to have a better representation in the predicted models during the evaluations, evaluate the results with statistical tests with respect to other approaches, investigate other regularization techniques as well as evaluate the proposed method in other datasets.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the National Council for Scientific and Technological Development CNPq (Grants 311404/2021-9 and 313643/2021-0) and the State of Minas Gerais Research Foundation - FAPEMIG (Grants APQ-00578-18 and APQ-01129-21).

REFERENCES

[1] L. Lowry and D. Linch, "Non-Hodgkin's lymphoma," *Medicine*, vol. 41, no. 5, pp. 282–289, 2013.

[2] T. A. A. Tosta, P. R. de Faria, V. R. Batista, L. A. Neves, and M. Z. do Nascimento, "Using Wavelet sub-band and Fuzzy 2-partition entropy to segment Chronic Lmphocytic Leukemia images," *Appl. Soft Comput.*, vol. 64, pp. 49–58, 2018.

[3] M. Sharif, M. Attique Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, no. 4, pp. 577–599, 2021.

[4] D. Gu, G. Liu, and Z. Xue, "On the performance of lung nodule detection, segmentation and classification," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101886, 2021.

[5] M. Ivanovici, N. Richard, and H. Decean, "Fractal dimension and lacunarity of psoriatic lesions-a colour approach," *medicine*, vol. 6, no. 4, p. 7, 2009.

[6] E. Zanaty and A. Afifi, "Generalized Hermite kernel function for Support Vector Machine classifications," *International Journal of Computers and Applications*, vol. 42, no. 8, pp. 765–773, 2020.

[7] L. C. Padierna, M. Carpio, A. Rojas-Domínguez, H. Puga, and H. Fraire, "A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: The Gegenbauer family," *Pattern Recognition*, vol. 84, pp. 211–225, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320318302280

[8] A. S. Martins, L. A. Neves, P. R. Faria, T. A. Tosta, D. O. Bruno, L. C. Longo, and M. Z. do Nascimento, "Colour feature extraction and polynomial algorithm for classification of lymphoma images," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2019, pp. 262–271.

[9] S. Wu, H. Jiang, H. Shen, and Z. Yang, "Gene selection in cancer classification using sparse Logistic Regression with l1/2 regularization," *Applied Sciences*, vol. 8, no. 9, p. 1569, 2018.

[10] D. Candelero, G. F. Roberto, M. Z. Do Nascimento, G. B. Rozendo, and L. A. Neves, "Selection of CNN, Haralick and Fractal features based on evolutionary algorithms for classification of histological images," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020, pp. 2709–2716.

[11] M. G. Ribeiro, L. A. Neves, M. Z. do Nascimento, G. F. Roberto, A. S. Martins, and T. A. A. Tosta, "Classification of colorectal cancer based on the association of multidimensional and multiresolution features," *Expert Systems with Applications*, vol. 120, pp. 262–278, 2019.

[12] E. Mortaz, "Imbalance accuracy metric for model selection in multi-class imbalance classification problems," *Knowledge-Based Systems*, vol. 210, p. 106490, 2020.

[13] L. Shamir, N. Orlov, D. M. Eckley, T. J. Macura, and I. G. Goldberg, "Iicbu 2008: a proposed benchmark suite for biological image analysis," *Medical & biological engineering & computing*, vol. 46, no. 9, pp. 943–947, 2008.

[14] A. S. Mubarak, S. Serte, F. Al-Turjman, Z. S. Ameen, and M. Ozsoz, "Local binary pattern and deep learning feature extraction fusion for COVID-19 detection on computed tomography images," *Expert Systems*, 2021.

[15] A. S. Martins, L. A. Neves, P. R. de Faria, T. A. Tosta, L. C. Longo, A. B. Silva, G. F. Roberto, and M. Z. do Nascimento, "A Hermite polynomial algorithm for detection of lesions in lymphoma images," *Pattern Analysis and Applications*, vol. 24, no. 2, pp. 523–535, 2021.

[16] R. F. Voss, "Random fractals: Characterization and measurement," in *Scaling phenomena in disordered systems*. Springer, 1991, pp. 1–11.

[17] H. Rhys, *Machine Learning with R, the tidyverse, and mlr*. Simon and Schuster, 2020.

[18] W. N. van Wieringen, "Lecture notes on Ridge regression," *arXiv preprint arXiv:1509.09169*, 2021.

[19] T. Shanableh and K. Assaleh, "Feature modeling using polynomial classifiers and stepwise regression," *Neurocomputing*, vol. 73, no. 10-12, pp. 1752–1759, 2010.

[20] V. Hooshmand Moghaddam and J. Hamidzadeh, "New Hermite orthogonal polynomial kernel and combined kernels in Support Vector Machine classifier," *Pattern Recognition*, vol. 60, pp. 921–935, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320316301534

[21] T. S. Chihara, *An introduction to orthogonal polynomials*. Courier Corporation, 2011.

[22] S. Thangavelu, "Hermite and laguerre semigroups: Some recent developments," in *Seminaires et Congres (to appear)*, 2006.

[23] L. A. Neves, M. Nascimento, D. Oliveira, A. S. Martins, M. Godoy, P. Arruda, D. de Santi Neto, and J. M. Machado, "Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5017–5029, 2014.

[24] H. Jiang, Z. Li, S. Li, and F. Zhou, "An effective multi-classification method for NHL pathological images," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018, pp. 763–768.

[25] N. Codella, M. Moradi, M. Matasar, T. Sveda-Mahmood, and J. R. Smith, "Lymphoma diagnosis in histopathology using a multi-stage visual learning approach," in *Medical Imaging 2016: Digital Pathology*, vol. 9791. International Society for Optics and Photonics, 2016, p. 97910H.

[26] G. F. Roberto, L. A. Neves, M. Z. Nascimento, T. A. Tosta, L. C. Longo, A. S. Martins, and P. R. Faria, "Features based on the percolation theory for quantification of Non-Hodgkin Lymphomas," *Computers in Biology and Medicine*, vol. 91, no. Supplement C, pp. 135 – 147, 2017.