Handcrafted features vs deep-learned features: Hermite Polynomial Classification of Liver Images

Handcrafted features vs deep-learned features: Hermite Polynomial Classification of Liver Images

Danilo C. Pereira* §, Leonardo C. Longo†, Thaína A. A. Tosta‡, Alessandro S. Martins§, Adriano B. Silva*,
Guilherme B. Rozendo†, Guilherme F. Roberto ¶, Alessandra Lumini ‖, Leandro A. Neves†, Marcelo Z. do Nascimento*
* Faculty of Computer Science (FACOM), Federal University of Uberlândia (UFU), Brazil
†Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP), Brazil
‡Science and Technology Institute, Federal University of São Paulo (UNIFESP), Brazil
§Federal Institute of Triângulo Mineiro (IFTM), Brazil
¶ Faculty of Engineering, University of Porto (FEUP), Portugal
‖Department of Computer Science and Engineering (DISI) - University of Bologna, Italy
E-mail: *danilo.pereira@ufu.br*

***Abstract*—Liver cancer is one of the most common types of cancer according to World Health Statistics. Computer-aided diagnosis (CAD) systems are used in medical imaging for liver tumor identification and classification. Texture is a type of feature that can provide measurements of properties such as smoothness and regularity of the image. Handcraft techniques based on fractal geometry allow quantifying self-similarity properties present in images. However, new studies have shown that using information obtained from deep-learned feature maps can maximize the results of classical classifiers. This work presents an approach that investigates descriptors obtained by handcrafted and deep learning features, feature selection methods and the Hermite polynomial (HP) algorithm to classifier liver histological images. The results were evaluated using metrics such as accuracy (ACC) and the imbalance accuracy metric (IAM). The association with fractal features, Lasso regularization and the HP algorithm achieved 0.98 of IAM and 99.53% ACC, which was relevant when evaluated with other studies in the literature.**

***Keywords*—Liver Tissue, Hermite Polynomial, Handcrafted features, Deep-learned Features, Feature Selection.**

I. Introduction

Pattern recognition and classification have gained popularity in recent years, particularly in the field of medical imaging in order to develop computer-aided diagnosis (CAD) systems. These tools are employed as supplementary reading in these tasks and include the evaluation of the digitized image, preprocessing, segmentation, feature extraction and classification. The feature extraction and selection modules are crucial in improving the classification performance allowing reduce computational complexity and decreasing dimensionality [1].

Liver cancer (LC) is the sixth most common type of cancer and the third leading cause of death from cancer worldwide. In the year 2022, 960,000 diagnoses of LC were recorded, accounting for 830,000 deaths [2]. Excessive alcohol use, smoking, family history, diabetes, obesity, hepatitis B or C virus infection and low immunity are factors that lead to LC [3]. In recent years, a significant number of researchers have been working on the development of the stages of a CAD system for the analysis of lesions, such as tumors present in the liver [4], [5].

The choice of the most relevant features is directly related to the image quality, the method used in feature extraction steps and the classification algorithm. A variety of feature extraction techniques can influence the performance of classifiers in CAD systems, such as overtraining [6] or heuristic dependencies of a specific group of features [7]. Among the approaches, an important technique for description is the quantification of an image's texture. 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 images, which can not be defined by Euclidean geometry [8].

Convolutional neural networks (CNNs) are approaches that allow the extraction of image feature patterns, taking into account multiple observation scales. This is accomplished through deep layers that enable the quantification of global and local patterns, known as deep-learned features. Recent studies have shown that using information obtained from CNN feature maps can maximize the results of classical classifiers [9], [10]. However, there are certain real-world problem-solving situations where available training data is limited and large datasets do not exist. In these cases, applying deep learning methods is not a viable option, conventional feature extraction and classification techniques may be a suitable solution [11].

Machine learning (ML) classification is a task that employs features as the basis for assigning class labels to patterns obtained from the input images [12]. Among the ML methods, the polynomial classifier is an algorithm that has shown promising results, especially when dealing with non-linearly separable data. This algorithm is a parameterized method that exponentially expands its polynomial basis according to the number of elements in the data vector and the function degree. Polynomial algorithms have demonstrated relevant performance in the process of analysis and labeling of samples during the classification stage. Among the bases, Hermite polynomials (HP) are able to generate a complete orthogonal basis of the Hilbert space that satisfy the orthogonality and completeness conditions of that space's family of elements.

The studies presented by [13], [14] explored only information in a binary approach and evaluated only part of handcrafted representations. The definition of suitable features or information has motivated new investigations to evaluate handcrafted and deep-learned features for describing important features, as well as analyzing the behavior of HP for multiclass classification with this type of information, which is still a challenging task in CAD systems.

The association of feature extraction and selection techniques with classifiers is a strategy that can be relevant and promising in a CAD system for age of the tissue classification. This study presents an approach based on descriptors obtained by handcrafted and deep-learned features, feature selection methods, and an HP classifier for the classification of liver histological images. The contributions of this work are:

- Study of the polynomial algorithm based on the multiclass Hermite basis function with a new parameterization for liver tumor classification;
- Investigation of Lasso regularization and Relief-F technique on descriptors obtained by handcrafted features and deep learning features;
- Performance comparison of the HP algorithm with ML methods and other studies in the literature.

II. Methodology

The proposed work performs the classification of liver tumor images through algorithms developed with 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 database employed in this work consists of images of female mouse liver tissue (Liver Age - LA) created by the Atlas of Gene Expression in Mouse Aging project (AGEMAP) [15]. All images in the database were digitized at a resolution of 417x312 pixels using a Carl Zeiss Axiovert 200 microscope and a 40x objective magnification. The total of 529 digitized images from different mice was divided into four classes, each of them representing a different age group of female mice on ad-libitum diets: one month (100), six months (115), 16 months (162), and 24 months (152). Fig. 2 shows examples of each class of digitized images.

B. Feature Extraction

The first stage of the methodology consists of feature extraction, which represents the transformations applied to image pixels to generate numerical values that are relevant for pattern discrimination [16]. Fractal dimension, lacunarity and percolation were computed, as demonstrated in [17]. 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 the Euclidean geometry [8].

The features of fractal dimension and lacunarity were obtained from probability matrices. In this work, these metrics were calculated using the gliding box method with a box size between $L = 3$ and $L = 43$ with an increment of two units at each iteration based in [17]. The probability matrix construction from these parameters ensured quantification at 20 different scales. Percolation metrics were extracted through the gliding-box algorithm, following the strategy described in [18]. Considering a box of size r, pixels in the box were considered pores and labeled with the application of the Hoshen-Kopelm algorithm [19]. Pores with the same label were considered part of the same cluster. This process was repeated for each box of side r [17]. Lacunarity and percolation were also represented by scalars in order to obtain representative descriptors of possible patterns in each curve. The scalar descriptors calculated were the maximum value, skewness, area under the curve and average between the halves of the area under the curve, as described in [18]. The feature vectors obtained with fractal geometry consisted of 116 descriptors.

Feature extraction was also performed using the CNN architectures ResNet-50 [20] and VGG-19 [21]. These CNNs were chosen due to their wide use in the literature and aim to provide a broader investigation of the proposed methods for classification and regularization based on features obtained through deep learning. From the ResNet-50, we targeted the final average pooling layer, which outputs 2048 features and, from VGG-19, 4096 features were obtained from the last fully-connected layer.

Each training considered 10 epochs, using the stochastic gradient descent (SGDM) strategy; initial learning rate $lr = 0.01$ with a reduction factor of 0.75 every 2 epochs; the cross-entropy function was used to calculate the adjustment on the parameters. This was repeated for each permutation of architecture and dataset. It is important to underline that the input images were normalized considering the standard deviation and average of the ImageNet dataset color channel values to match the methodology used on the model's pre-training [22]. Finally, the resulting fine-tuned models were the ones that achieved the highest accuracy for the evaluation set, independent of the epoch.

C. Lasso

Lasso is a regularization technique that aims to normalize the features by introducing terms that promote resizing and enhance the stability of the problem. This approach can address the issue of having more features than classes by allowing some input variables to be set to zero, resulting in a more efficient predictive model [23].

The smoothing of the objective function F is performed through Lasso regularization and can be achieved by introducing a penalty term equal to the absolute value of the coefficient β into the cost function, as shown in Eq. 1. The hyperparameter λ ($0 < \lambda < \infty$) controls the degree of regularization applied to each feature, penalizing them with varying weights. This regularization method can generate sparse models with only a few coefficients. In some cases, coefficients can even be reduced to zero and excluded from

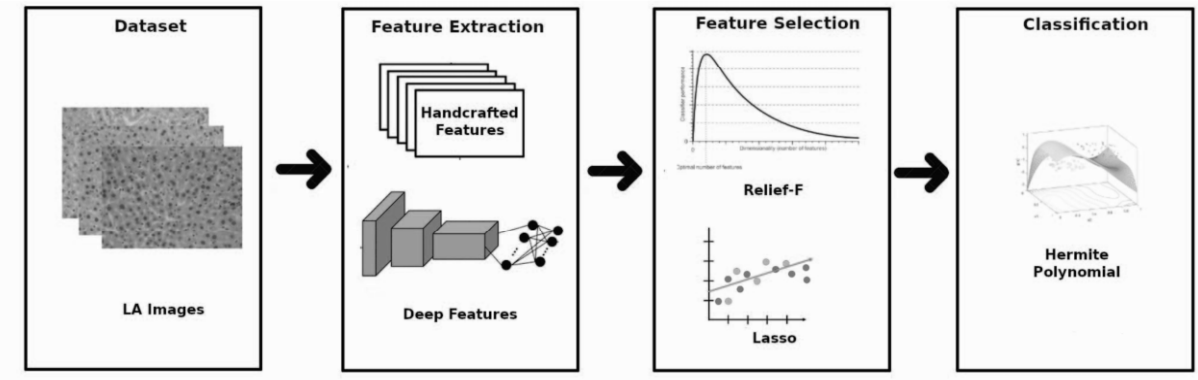

Fig. 1. Block diagram of the main stages of the proposed approach.

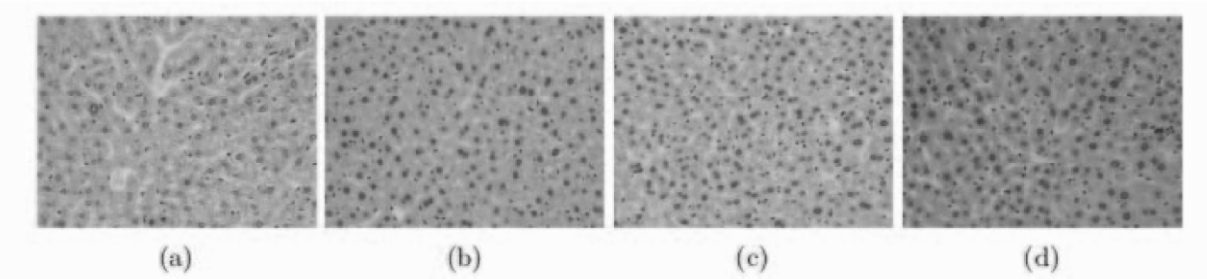

Fig. 2. Female mouse liver tissue samples from the LA database with the following ages: (a) 1, (b) 6, (c) 16 and (d) 24 months. Source: [**?**]

the model entirely. Larger penalties lead to coefficient values closer to zero, promoting the creation of simpler models.

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

The optimal value of λ, which produces the highest classification rate, can be determined iteratively using a cross-validation procedure. In this study, the range of λ values explored was between 10^{-6} and 10^{-2} with a total of $3{\times}10^{6}$ iterations. These values were chosen based on an empirical evaluation of the feature vectors. During the training process, the regularization stage was implemented in each fold to identify the most relevant features for classification. This approach reduced the number of features required for the classification stage [24]. At this stage, the features were ordered according to their importance weights and, then, four groups were obtained by selecting the 10, 15, 20 and 25 features. The selected features were applied to all of the classifiers investigated in this study.

D. Relief-F

The Relief-F method is a feature weighting technique that analyzes the dataset information and assigns weight values according to the association between each feature and the class. This technique selects the features based on the weight values that help distinguish instances that are close to each other [25]. According to [26], Relief-F has the ability to reduce the dimension of the feature set by removing negative values, which can maximize the efficiency of algorithms that perform multiclass classification. The Relief-F method also presents the ability to handle noisy data, efficient processing time and high accuracy.

For this method, Eq. 2 presents the operation of the feature selection. In this equation, $f_{t,i}$ describes the value of the instance x_i for the feature f_i, P represents the distance measure and the terms $f_{dc(x_i)}$ and $f_{sc(x_i)}$ are the values of the points of the *i*-th neighboring features of x_i with different or equal labels [27].

$$f_i = \frac{1}{2} \sum_{i=1}^{N} P(f_{t,i} - f_{dc(x_i)}) - P(f_{t,i} - f_{sc(x_i)}) \quad (2)$$

Similarly to the regularization method, vectors constructed in this algorithm were composed of 10, 15, 20 and 25 features, according to the importance level assigned to them with this method.

E. Classification

A polynomial classifier is a supervised approach that expands the input feature vector $\mathbf{x} = [x_1 \dots x_d]^T$, where d represents the number of features and T is the transpose operation, to a higher dimension in a nonlinear manner. This technique enables the generation of linear approximations in

this space, which can be used to classify the input data into the desired output [28].

In [29], Thangavelu describes that the HP algorithm is orthogonal on the interval $(-\infty, \infty)$ with respect to the standard Gaussian weight function and provides advantages in function approximation. The HP algorithm can be defined mathematically as shown in Eq. 3:

$$HP_n(\mathbf{x}) = (-1)^n e^{+\mathbf{x}^2/2} \frac{d^n}{d\mathbf{x}^n}[e^{-\mathbf{x}^2/2}]. \quad (3)$$

As a result, the HP method can be computed using the recurrence relation for any order where $n > 0$ (see Eq.4):

$$HP_{n+1}(\mathbf{x}) = \mathbf{x}HP_n(\mathbf{x}) - nHP_{n-1}(\mathbf{x}) = 0, \quad (4)$$

where $HP_0(\mathbf{x}) = 1$ and $HP_1(\mathbf{x}) = \mathbf{x}$.

Using the recurrence equation, the polynomials among the degree 1 to 3 are obtained by Eqs. (5), (6), and (7):

$$HP_2(\mathbf{x}) = \mathbf{x}^2 - 1, \quad (5)$$

$$HP_3(\mathbf{x}) = \mathbf{x}^3 - 3\mathbf{x}, \quad (6)$$

$$HP_4(\mathbf{x}) = \mathbf{x}^4 - 6\mathbf{x}^2 + 3. \quad (7)$$

For the classification of image groups, the feature vectors defined by **x** were used as inputs and then expanded in terms of the polynomial basis $HP_n(\mathbf{x})$.

For the multiclass classification, the decision rule can be expressed using Eq. 8. In this equation, the *i*-th problem is addressed by a linear discriminant function that separates the points assigned to w_i from those not assigned to w_i. The decision surface, given by $g(\mathbf{x}) = 0$, demarcates the boundary between the classes w_i and $not(w_i)$ for a multiclass problem.

$$\text{Decide} \begin{cases} \omega_i, & \text{if} \quad g(\mathbf{x}) > 0 \\ not(\omega_i), & \text{if} \quad g(\mathbf{x}) < 0 \end{cases} \quad (8)$$

Finally, the output $g(\mathbf{x})$ can be obtained by Eq. 9.

$$g(\mathbf{x}) = \mathbf{a}^T HP_n(\mathbf{x}), \quad (9)$$

where the coefficient vector of the polynomial basis function denoted by **a**, $HP_n(\mathbf{x})$ represents the *Hermite* basis function, and n corresponds to the order or degree of the polynomial. This algorithm was divided into two steps, namely training and testing, which are detailed in [30].

In this work, the degree of the polynomial expansion was determined empirically, where the relevant results were achieved using a third-order polynomial for class separation.

F. Evaluation of Methods

The HP algorithm was compared with three different classifiers based on the primary supervised ML approaches, namely function-based, ensemble learning, and tree-based. The selected algorithms were logistic regression (LGT), which is a model that integrates tree induction and additive logistic regression; multilayer perceptron (MLP), an approach that employs a system of interconnected neurons or nodes to map nonlinear relationships between input and output vectors; and random forest (RF), a strategy that utilizes a tree ensemble method, with bootstrapping for each tree generating subsets of observations not included in the tree-growing process. These algorithms and HP classifier were evaluated using the cross-validation method with *k*=10, with 90% of the dataset used for training and 10% for testing the model.

The IAM and ACC were used to evaluate the classification algorithms. ACC is a widely used metric in image classification analysis because it is easy to calculate and interpret, and ranges between 0 and 100%. A perfect classification results in an accuracy of 100%. However, ACC may not be a reliable measure for unbalanced class problems. The IAM metric, defined by Eq. 10, was also applied in the experiments to address the issue of unbalanced labels and improve the recall of results based on the data features.

$$IAM = 1/k \sum_{i=1}^{k} \frac{c_{ii} - max(\sum_{j \neq i}^{k} c_{ij}, \sum_{j \neq i}^{k} c_{ji})}{max(c_{.i}, c_{i.})}, \quad (10)$$

where 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 achieved with the IAM and ACC metrics for the HP classifier associated with the Lasso regularization considering a range of five features obtained of range among 10 to 25. This table presents the sets of features obtained with three approaches, one using handcrafted techniques and the other with deep-learned features. Analyzing the data, it can be seen that the HP algorithm and its associations showed more relevant results with the handcrafted descriptor using fractal techniques and 25 features, achieving an ACC of 99.53% and an IAM value of 0.98. Regarding the descriptors obtained with the ResNet-50 model, the results were 89.87% for ACC and 0.49 for the IAM metric. The VGG-19 model presented 91.47% and 0.57 for the ACC and IAM metrics, respectively, with 25 features. In this case, the use of 20 features also showed similar results for the ACC metric, but on IAM, the performance with 25 features was better, including a lower standard deviation. IAM is a more robust metric for evaluating classifiers on multiclass datasets, which offers benefits in imbalanced data. It is noticed in these experiments that the fractal descriptors were more robust regarding this metric, as the performances ranged from 0.86 to 0.98. When applied to the descriptors from CNNs, this metric was lower in all experiments, despite the number of features used. When applied to the descriptors from CNNs, this metric was lower in all experiments, despite the number of features used. IAM has a representation range between -1 to 1 and, when this value is closer to zero, it indicates that the number of correctly and incorrectly classified instances is close. As mentioned by the authors in [31], the IAM metric allows for a better evaluation

TABLE I
RESULTS OBTAINED WITH THE HP CLASSIFICATION ALGORITHM AND LASSO REGULARIZATION.

Data	Number	ACC (%)	IAM
Fractal	10	97.44±0.15	0.86±0.09
	15	98.49±0.14	0.91±0.07
	20	99.24±0.09	0.95±0.06
	25	**99.53±0.08**	**0.98±0.03**
ResNet-50	10	80.88±0.03	0.12±0.12
	15	85.05±0.03	0.28±0.13
	20	87.88±0.03	0.40±0.13
	25	89.87±0.01	0.49±0.07
VGG-19	10	85.32±0.03	0.30±0.10
	15	89.58±0.02	0.49±0.10
	20	91.28±0.02	0.56±0.11
	25	91.47±0.02	0.57±0.10

of the classifier's behavior since it provides information about mislabeling instance classes.

The results achieved by the HP algorithm in association with the Relief-F selector are presented in Table II for intervals between 10 and 25 features. Upon evaluation, it is noted that the HP algorithm provided the best values using fractal descriptors with 25 features, ACC of 99.43% and IAM of 0.96. The same behavior occurred with the deep-learned descriptors, using ResNet-50 and VGG-19, where 25 characteristics with Relief-F allowed ACC values of 98.39% and 97.44%, respectively. With the IAM metric, these descriptors using the Relief-F provided better results than those achieved with HP and Lasso regularization. However, these values were lower compared to those obtained with fractal descriptors (25 features), Relief-F and HP classifier. In addition, the ResNet-50 network allowed an IAM value of 0.88, considering the same feature number. VGG-19 achieved an IAM of 0.85 with the same number of features.

The results depicted in these tables showed that the association of the 25 most relevant fractal descriptors, Lasso regularization, and HP classifier allowed the achievement of more effective values. It is also possible to verify that the standard deviation values with handcrafted features were lower for the IAM metric, regardless of the feature selection technique.

Based on the results presented in Tables I and II, the association of the HP algorithm with the Lasso regularization (25 features) were compared with ML algorithms. Table III displays the results achieved by these strategies. The MLP and RF algorithms achieved values of 96.04% and 94.15% for ACC and 0.97 with the IAM metric for both methods. The LGT algorithm obtained 91.53% ACC and 0.96 for IAM. It is important to highlight that the standard deviation of all classifiers was below 0.05% for both metrics. According to the results, the ML algorithms had lower performance than the results with the HP algorithm and proposed associations.

The indirect analysis of the proposed association and other recent works that investigated computational techniques for the classification of LA images are presented in Table IV. It can be observed that the association of the HP algorithm, Lasso, and fractal descriptors showed promising results compared to other works available in the literature, indicating a relevant solution to assist experts in the analysis of this type of image.

TABLE II
RESULTS OBTAINED WITH HP CLASSIFICATION ALGORITHM WITH RELIEF-F SELECTOR EMPLOYED IN LA IMAGES.

Data	Number	ACC (%)	IAM
Fractal	10	93.27±0.01	0.64±0.06
	15	95.74±0.02	0.76±0.10
	20	98.67±0.01	0.91±0.05
	25	**99.43±0.01**	**0.96±0.04**
ResNet-50	10	94.13±0.02	0.70±0.09
	15	96.60±0.01	0.80±0.08
	20	97.64±0.01	0.84±0.06
	25	98.39±0.01	0.88±0.05
VGG-19	10	92.32±0.02	0.60±0.13
	15	95.17±0.01	0.73±0.08
	20	96.21±0.02	0.78±0.10
	25	97.44±0.02	0.85±0.10

TABLE III
CLASSIFICATION OBTAINED WITH THE FRACTAL DESCRIPTOR, LASSO REGULARIZATION AND ML ALGORITHM.

Classifiers	Number	ACC (%)	IAM
LGT	25	91.53±0.04	0.96±0.02
MLP	**25**	**96.04±0.01**	**0.97±0.01**
RF	25	94.15±0.02	0.97±0.01

TABLE IV
ANALYSIS OF THE ACCURACY METRIC FROM DIFFERENT APPROACHES DEVELOPED IN THE LITERATURE.

Reference	Approach	ACC (%)
Roberto et al. [5]	ResNet-50, FD, LAC and PERC (DLHC)	99.62
Proposed Method	Handcrafted Features Lasso and HP Classifier	99.53
Andrearczyk et al. [32]	Collective T-CNN	98.20
Huang et al. [33]	Novel set of image features and Ensemble SVM Classifier	97.01
Watanabe et al. [4]	GIST descriptors, PCA and LDA (HC)	88.40

IV. CONCLUSION

The present study introduced a computational tool for analyzing LA images based on fractal descriptors, Lasso regularization, and the HP classifier. The analyses presented in this study explored the association between feature selectors and the HP algorithm for building prediction and classification models of LA histological images. The obtained results indicated that the proposed approach using a set of 25 handcrafted features through the Lasso regularizer presented the best performance, with values of IAM and ACC metrics exceeding 0.98 and 99.53%, respectively.

Values obtained with the CNN-based descriptors were more relevant when the Relief-F selector was applied for the ACC metric. The same behavior is observed with the IAM metric. In

future studies, it is intended to analyze which features provide the most satisfactory results, as well as to measure the gain in relation to the computational cost of the HP algorithm when performing the training and testing stages using the proposed approach. Further, the number of mice is low and images from the same animal can be used in the training and test sets (16 different animals). However, we can find great variability in the cells of the same mice. In addition, we plan to explore the feature selection techniques and evaluate the effectiveness of the proposed method in other histological image dataset.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of National Council for Scientific and Technological Development CNPq (Grants #313643/2021-0, #311404/2021-9), the State of Minas Gerais Research Foundation - FAPEMIG (Grant #APQ-00578-18 and Grant #APQ-01129-21) and São Paulo Research Foundation - FAPESP (Grant #2022/03020-1).

REFERENCES

[1] E. Tasci and A. Ugur, "A novel pattern recognition framework based on ensemble of handcrafted features on images," *Multimedia Tools and Applications*, vol. 81, no. 21, pp. 30 195–30 218, 2022.

[2] U. Cinar, R. Cetin Atalay, and Y. Y. Cetin, "Human hepatocellular carcinoma classification from h&e stained histopathology images with 3d convolutional neural networks and focal loss function," *Journal of Imaging*, vol. 9, no. 2, p. 25, 2023.

[3] K. M. Napte and A. Mahajan, "Liver segmentation using marker controlled watershed transform," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, p. 1541, 2023.

[4] K. Watanabe, T. Kobayashi, and T. Wada, "Semi-supervised feature transformation for tissue image classification," *Plos one*, vol. 11, no. 12, p. e0166413, 2016.

[5] G. F. Roberto, A. Lumini, L. A. Neves, and M. Z. do Nascimento, "Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images," *Expert Systems with Applications*, vol. 166, p. 114103, 2021.

[6] S. Tripathi and S. K. Singh, "Ensembling handcrafted features with deep features: an analytical study for classification of routine colon cancer histopathological nuclei images," *MULTIMEDIA TOOLS AND APPLICATIONS*, 2020.

[7] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.

[8] 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.

[9] E. F. Ohata, J. V. S. d. Chagas, G. M. Bezerra, M. M. Hassan, V. H. C. de Albuquerque *et al.*, "A novel transfer learning approach for the classification of histological images of colorectal cancer," *The Journal of Supercomputing*, vol. 77, no. 9, pp. 9494–9519, 2021.

[10] N. Kumar, M. Sharma, V. P. Singh, C. Madan, and S. Mehandia, "An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images," *Biomedical Signal Processing and Control*, vol. 75, p. 103596, 2022.

[11] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Handcrafted versus cnn features for ear recognition," *Symmetry*, vol. 11, no. 12, 2019. [Online]. Available: https://www.mdpi.com/2073-8994/11/12/1493

[12] 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.

[13] 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.

[14] D. C. Pereira, L. C. Longo, T. A. Tosta, A. S. Martins, A. B. Silva, P. R. de Faria, L. A. Neves, and M. Z. Do Nascimento, "Classification of lymphomas images with polynomial strategy: An application with ridge regularization," in *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, vol. 1. IEEE, 2022, pp. 258–263.

[15] J. M. Zahn, S. Poosala, A. B. Owen, D. K. Ingram, A. Lustig, A. Carter, A. T. Weeraratna, D. D. Taub, M. Gorospe, K. Mazan-Mamczarz *et al.*, "Agemap: a gene expression database for aging in mice," *PLoS genetics*, vol. 3, no. 11, p. e201, 2007.

[16] 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.

[17] 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.

[18] 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.

[19] J. Hoshen and R. Kopelman, "Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm," *Physical Review B*, vol. 14, no. 8, p. 3438, 1976.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] "torchvision.models," Jun 2021. [Online]. Available: https://pytorch.org/vision/stable/models.html

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

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

[25] K. Liu, Q. Chen, and G.-H. Huang, "An efficient feature selection algorithm for gene families using nmf and relieff," *Genes*, vol. 14, no. 2, p. 421, 2023.

[26] Y. M. Yacob, H. Alquran, W. A. Mustafa, M. Alsalatie, H. A. M. Sakim, and M. S. Lola, "H. pylori related atrophic gastritis detection using enhanced convolution neural network (cnn) learner," *Diagnostics*, vol. 13, no. 3, p. 336, 2023.

[27] L. O. Felix, D. H. C. de Sá Só, U. A. B. V. Monteiro, B. M. Castro, L. A. V. Pinto, C. A. O. Martins *et al.*, "A feature selection committee method using empirical mode decomposition for multiple fault classification in a wind turbine gearbox," 2023.

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

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

[30] 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, pp. 523–535, 2021.

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

[32] V. Andrearczyk and P. F. Whelan, "Deep learning for biomedical texture image analysis," in *Proceedings of the Irish Machine Vision & Image Processing Conference*. Irish Pattern Recognition & Classification Society (IPRCS), 2017.

[33] H.-L. Huang, M.-H. Hsu, H.-C. Lee, P. Charoenkwan, S.-J. Ho, and S.-Y. Ho, "Prediction of mouse senescence from he-stain liver images using an ensemble svm classifier," in *Intelligent Information and Database Systems: 5th Asian Conference, ACIIDS 2013, Kuala Lumpur, Malaysia, March 18-20, 2013, Proceedings, Part II 5*. Springer, 2013, pp. 325–334.