ORIGINAL ARTICLE

Grading the Dominant Pathological Indices in Liver Diseases from Pathological Images Using Radiomics Methods

Hamed Zamanian, Ahmad Shalbaf *

Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

*Corresponding Author: Ahmad Shalbaf
Email: shalbaf@sbmu.ac.ir

Received: 10 December 2022 / Accepted: 05 February 2023

Abstract

Purpose: This study aims to diagnose the severity of important pathological indices, i.e., fibrosis, steatosis, lobular inflammation, and ballooning from the pathological images of the liver tissue based on extracted features by radiomics methods.

Materials and Methods: This research uses the pathological images obtained from liver tissue samples for 258 laboratory mice. After preprocessing the images and data augmentation, a collection of texture feature sets extracted by gray-level-based algorithms, including Global, Gray-level Co-Occurrence Matrix (GLCM), Gray-level Run length Matrix (GLRLM), Gray-level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM) algorithms. Then, advanced methods of classification, namely Support Vector Machine (SVM), Random Forest (RF), Quadratic Discriminant Analysis (QDA), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naïve Bayes (NB), and Multi-layer Perceptrons (MLP) are employed. This procedure is provided separately for each of the four indices of fibrosis level in 6 grading classes, steatosis in 5 grading classes, inflammation in 4 grading classes, and ballooning in 3 grading classes. For a comparison of the output of these algorithms, the accuracy value obtained from the evaluation data is presented for the performance of different methods.

Results: The results showed that, compared to other methods, the Gaussian SVM algorithm provides a better response to the classification of the grading of liver disease among all the indices from the pathological images due to its structural features. This value of accuracy was calculated at 84.30% for fibrosis, 90.55% for steatosis, 81.11% for inflammation, and 95.98% for ballooning.

Conclusion: This fully automatic framework based on advanced radiomics algorithms and machine learning from pathological images can be very useful in clinical procedures and be considered as an assistant or a substitute for pathologists' diagnoses.

Keywords: Liver Disease; Machine Learning; Radiomics; Gaussian Support Vector Machine; Pathological Images.

DOI:

1. Introduction

Non-alcoholic fatty liver disease is one of the most important diseases in the world and occurs in people with no history of alcohol addiction [1]. One of the most common types of non-alcoholic fatty liver diseases is getting hepatitis based on non-alcoholic fat or simpler, Non-alcoholic Steatohepatitis (NASH). In this disease, the extra fat made in the liver causes inflammation in the liver tissue. This disease in more severe conditions can be associated with liver tissue scarring and eventually liver cirrhosis [2]. Today, four variables indices named steatosis, fibrosis, inflammation, and ballooning determine the stage and the level of severity of various Non-alcoholic fatty liver diseases [3, 4]. Based on the values of these four indices, the NAFLD Activity Score (NAS) index is calculated as a standard acceptable measure by the public for the level of Non-alcoholic fatty liver diseases [3]. Steatosis is common for most grades of NASH [5]. Lobular inflammation is also commonly seen in different grades of NASH and is considered one of the prerequisites for their occurrence. Moreover, the level of fibrosis is different in various steps of NASH. Fibrosis occurs because of the severe destruction of liver tissue due to the accumulation of Extracellular matrix (ECM) proteins. In addition to fat accumulation in hepatocytes with inflammation in lobules, if the fibrosis is only portal or periportal of the liver, NASH is classified in stage 1. For central and portal fibrosis, NASH is classified in stage 2. For fibrosis with bridging between inflamed hepatocytes, NASH is defined in stage 3. Stage 4 of NASH will be diagnosed only in the condition of occurring liver cirrhosis [6].

To determine these indices, the samples collected from the liver tissue are taken with the help of biopsy needles or during the laparoscopic processes, and then the diagnosis process is done by the pathologist based on the liver pathological images. An experienced pathologist in the field of liver diseases can classify each of these indices from the pathological images collected from the liver sample. But performing this process by a pathologist is so time-consuming, tiring, dependent on the person's opinion, and requires a high level of expertise and experience. The complexity and entanglement of the occurrence of each of the pathological changes that are associated with the appearance in the pathological images make it difficult to accurately diagnose each index. Therefore, designing a new and fully automatic framework based on advanced machine learning algorithms can be very useful and helpful to recognize this process from pathological images in clinical procedures. Because they can recognize complex patterns in large data sets. These intelligent algorithms can be considered an assistant or a substitute for pathologists' diagnoses.

Different approaches of machine learning including deep learning and radiomics algorithms have been used in liver diseases [7, 8]. Weiner introduced an approach based on artificial intelligence to assess the sample of liver pathology to determine the severity of the disease and its heterogeneity accurately [9]. Roy and his group introduced a network based on deep learning to exactly quantify steatosis based on pathology images [10]. In this work, an area is extracted and the border between the steatotic area and the background is determined. This assessment provides a very good correlation with the pathologist's result. Jana's group also used a combined method with the information obtained from pathology and CT images to improve the estimation of fibrosis degree and NAS score [11]. They jointly applied CT scans and pathology images as input to Convolutional Neural Network (CNN) and investigated the effect of different formats of these data applications. The report of Arjmand's group is also based on a comprehensive review of a diagnostic process with the help of deep learning CNN network from pathological images obtained from liver biopsy [12]. Their main goal is to provide a relatively accurate diagnostic mechanism for classifying the degree of ballooning and fat accumulation in sample tissue. These two indices are finally divided into four separate classes with the help of two CNN networks and two different optimization functions. Forlano also measures the quantitative percentage of pathological indices with the help of machine learning according to the result of the pathologist's experimental evaluation [13]. This approach helps to identify the difference between the investigated area with higher sensitivity, in comparison with the common grading system. In 2019, Yang utilized a machine learning algorithm to estimate the value of steatosis from pathological images [14]. For this purpose, some sections are separated and labeled from the original image. Qu's group also used the division of the pathology image into smaller parts to locally evaluate the four pathological indices proposed for the NAS score as well as the degree of fibrosis, and finally obtain an overall conclusion [15]. Also, Yumeng used the textural and morphological features obtained from the pathological images to extract the grade of NAS [16]. He diagnosed

NASH/NAFLD with the help of 43 morphological and 109 textural features and followed it with various classification methods. Heinemann also tried to grade NASH models based on pathological images with the help of CNN learning networks [17]. They presented a classification approach based on targeted indices to achieve the NAS score and the severity of the NASH disease using the collected liver samples. In their study, four different classification networks were designed, and the networks were trained and tested based on the opinion of an experienced pathologist about the class of images. The deep learning CNN used in this work was the Inception-ResNet V3 network, which obtained acceptable results. Recently, Heinemann published a report on deep learning quantification of pathological indices to access the NASH progression on human liver biopsy samples [18]. This research includes 467 biopsy samples and the structure of the network was made of 4 independent CNN networks to extract the relative features and yield average class probabilities. Finally, their outputs were scored by artificial neural network models.

In this study, we aim to diagnose the severity of pathological indices, i.e., fibrosis (in degrees of 0 to 4 with negligible cases), steatosis (in degrees of 0 to 3 with negligible cases), lobular inflammation (in degrees of 0 to 2 with negligible cases), ballooning (in degrees of 0 and 1 with negligible cases) from the pathological images of the liver tissue. This work is done by the combination of the extracted features by radiomics methods, including Global, Gray-level Co-Occurrence Matrix (GLCM), Gray-level Run length Matrix (GLRLM), Gray-level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM) algorithms and finally different advanced machine learning classifier.

2. Materials and Methods

2.1. Dataset

The data sets used in this research are from the animal samples studied in [17], which were collected and analyzed in the period from 2014 to 2019. These animals are categorized into two groups, healthy controls and affected with NASH/NAFLD. The experiments on these animals, which included 258 mice (RjHan: WI and C57BL/6JRj) of different ages, were carried out after receiving the necessary approvals from Regierungsprasidium Tubingen, Germany. After removing the liver tissue from them, a 3-micrometer piece was separated from the central area of the right lobe and stained by Masson tri-chrome. To image from these separated pieces, a Zeiss AxioScan Z1 scanner [19] was employed with a 20x magnification based on bright background lighting, with a pixel resolution of 0.22µm/px. The assessment of images is provided based on categorizing the extracted segments. This segmentation has been done considering the lack of overlap and their classification has been done in different resolutions for 4 indices. In this way, low resolutions from pieces of images were extracted for fibrosis with dimensions of 299*299 px^2 and a resolution of 1.32µm/px. For the other three cases of ballooning, lobular inflammation, and steatosis, images with dimensions of 299*299px^2 with a pixel resolution of 0.44µm/px were used. The grading process of liver tissue pathology images for the four desired indices of fibrosis (in degrees of level 0 to 4 with negligible cases), steatosis (in degrees of level 0 to 3 with negligible cases), lobular inflammation (in degrees of Level 0 to 2 with negligible cases), ballooning (level grades 0 and 1 with negligible cases) was performed by an experienced liver pathologist. Figure 1 shows an example of different pathological indices in different grades. According to this, in fibrosis, 4368, in steatosis, 6377 segments, in lobular inflammation, 8701 segments, and in ballooning, 14465 segments have been extracted and used. In Table 1, the numbers of images in different classes of the four indices (Fibrosis, Steatosis, Inflammation, and Ballooning) are summarized.

2.2. An Overview of the Proposed Algorithm

The overall structure of the proposed method in this work is briefly shown in Figure 2. Pathological images are obtained from liver tissue samples for 250 laboratory mice. Then, preprocessing is done to balance the number of images of different classes. New pathological images are reproduced with the help of common simple algorithms of data augmentation such as rotating, reshaping, and cropping. Using this process of balancing the number of samples leads to the implementation of classification algorithms based on a balanced data bank, which minimizes the possibility of biasing the evaluation results on a specific class. This balancing was done simply by making initial changes to the existing data in each class and creating new

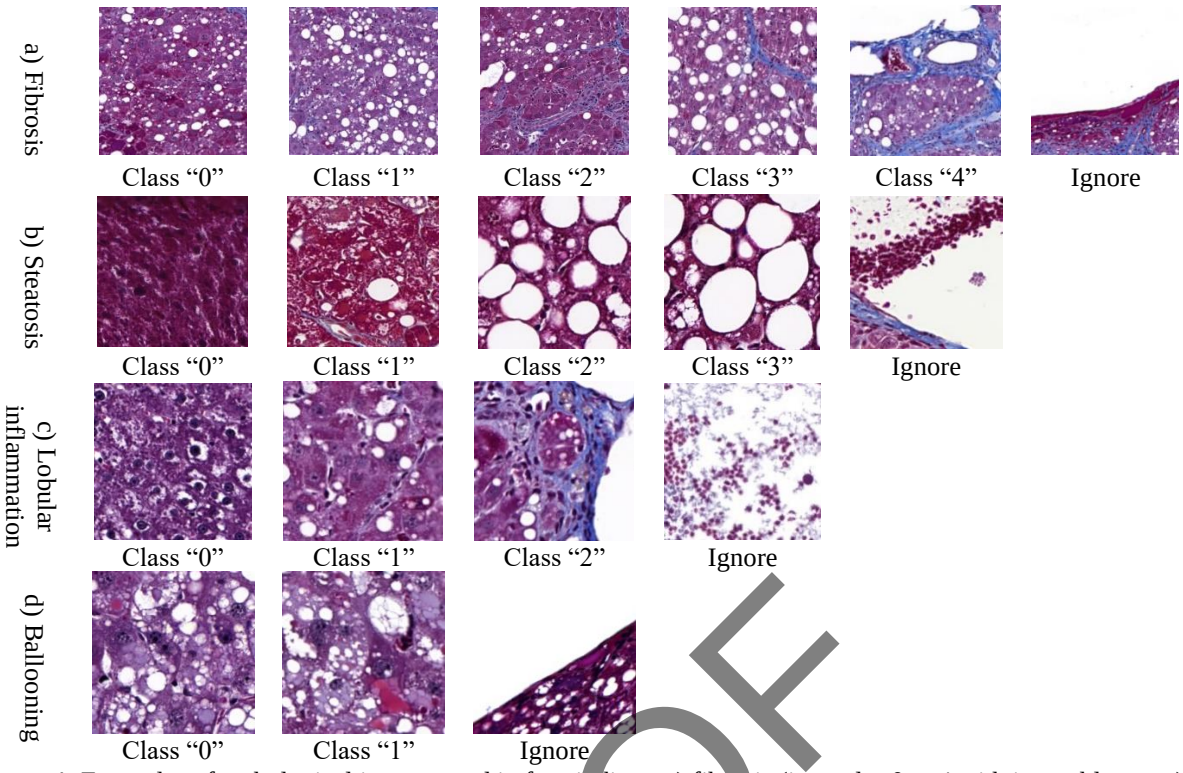

Figure 1. Examples of pathological images used in four indices: a) fibrosis (in grades 0 to 4 with ignorable cases), b) Steatosis (in grades 0 to 3 with ignorable cases), c) Lobular inflammation (in grades 0 to 2 with ignorable cases), and d) ballooning (in grades 0 and 1 with ignorable cases)

Table1. The number of images in different classes of the four indices (Fibrosis, Steatosis, Inflammation, and Ballooning) before and after augmentation

Index	No. of images before augmentation	No. of images after augmentation	Class 0	Class 1	Class 2	Class 3	Class 4	Ignore
Fibrosis	4368	14458	2299	2887	2419	2500	2209	2444
Steatosis	6377	10100	1175	1378	2683	2581	-	2283
Inflammation	8701	23767	5533	5736	6521	-	-	5977
Ballooning	14465	37065	11936	12948	-	-	-	12181

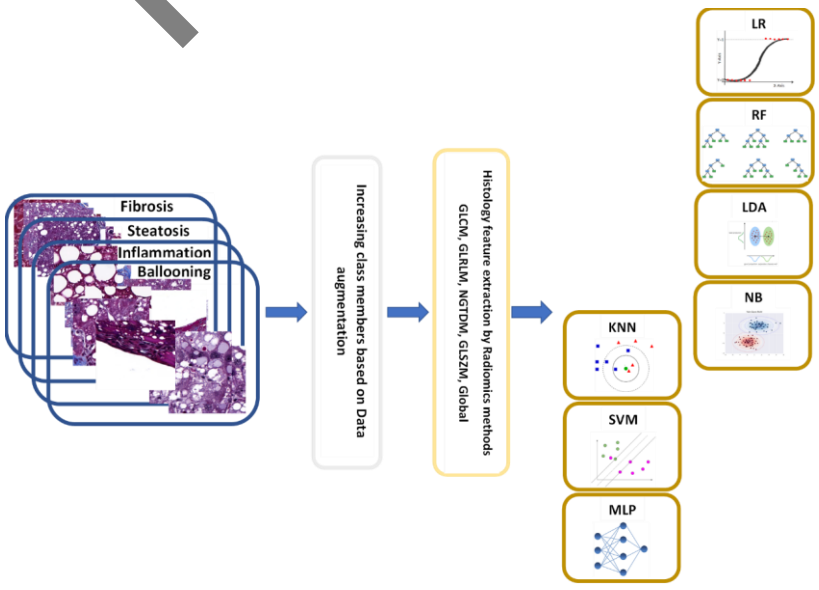

Figure 2. The overall structure of the proposed method for classifying pathological images in four indices (Fibrosis, steatosis, lobular inflammation, and ballooning)

samples without adding new information. Then, different texture features, including Global, GLCM, GLRLM, GLSZM, and NGTDM are extracted and after normalizing, they are fed to different classification methods, SVM, LR, RF, QDA, KNN, NB, and MLP to classify the data in the training and testing model. This procedure of classification of pathological images of liver samples has been done separately for each of the four indices of steatosis, fibrosis, lobular inflammation, and ballooning. In fibrosis, 6 different classes, in steatosis, 5 different classes, in lobular inflammation, 4 different classes, and in ballooning, 3 different classes are considered. The training and validation of the present algorithms were implemented with the help of MATLAB and Python software. All this process was done in the Windows platform using a system with a Ryzen 9 5900 Central Processing Unit (CPU), GeForce RTX 3060 Ti Graphical Processing Unit (GPU), and 16 gigabytes of RAM.

2.3. Pre-Processing

The collected images, for each of the pathological indices, have different numbers in their different grading that show an appropriate imbalance in members. This difference can cause a weighting bias for a class with more members in the process of training and as a result, the procedure of the test will be faced with errors in other classes. Therefore, it is necessary to use data augmentation algorithms. Data augmentation methods refer to algorithms that create artificial additional data based on real members of classes, without imposing new information on the model [20]. In other words, for grading classes with fewer data, a new data series with the general characteristics of the main members are reproduced to make the data of that class competitive with other classes. Different algorithms have been proposed to implement this approach which can be done in the simplest situation by changing the color, rotation angle, cropping, noise level, spatial shift, etc. In this work, for simplicity and describing better the final data, we limited the data augmentation only by rotation with a small degree, reshaping, and cropping. The number of members of classes with more members was considered a reference and the number of members of other classes is increased according to the difference with that class in 4 indices. According to this, the number of images in different classes of the four indices (Fibrosis, Steatosis, Inflammation, and Ballooning) after augmentation is presented in Table 1.

2.4. Extracting the Image Features

Different features related to the changes between the neighbor pixels in the pathological images are extracted by various approaches, including Global, Gray-level Co-Occurrence Matrix (GLCM), Gray-level Run length Matrix (GLRLM), Gray-level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). These sets of features are known as Radiomics features, which provide comprehensive information about the shape, brightness, and texture of the image. These extracted features are known as second and higher-order features that are determined by considering the relation between neighboring pixels. In the GLCM method, calculations are performed in four different directions with angle changes of 45 degrees at angles of 45, 90, 135, and 180 degrees, and for three different pixel steps resulting from one-, two-, and three-pixel neighborhoods. After that, the results from four different directions are averaged. In the Global method, 3 features are extracted: Variance, Skewness, and Kurtosis. In GLCM, the extracted features are Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Maximum probability, Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation1, Information measure of correlation 2, Inverse difference (INV) homogenous, Inverse difference normalized (INN), and Inverse difference moment normalized. In the GLRLM method, the extracted features are Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray-Level Non-uniformity (GLN), Run-Length Nonuniformity (RLN), Low Gray-Level Run Emphasis (LGRE), High Gray-Level Run Emphasis (HGRE), Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SRHGE), Long Run Low Gray-Level Emphasis (LRLGE), Long Run High Gray-Level Emphasis (LRHGE), and Gray-Level Variance (GLV). In NGTDM, the extracted features are Coarseness, Contrast, Busyness, Complexity, and Strength. Finally, the extracted features for GLSZM are as follows: Small Zone Emphasis (SZE), Large Zone Emphasis (LZE), Gray-Level Nonuniformity (GLN), Zone-Size Non-uniformity (ZSN), Zone Percentage (ZP), Low Gray-Level Zone Emphasis (LGZE), High Gray-Level Zone Emphasis (HGZE),

Small Zone Low Gray-Level Emphasis (SZLGE), Small Zone High Gray-Level Emphasis (SZHGE), Large Zone Low Gray-Level Emphasis (LZLGE), Large Zone High Gray-Level Emphasis (LZHGE), Gray-Level Variance (GLV), and Zone-Size Variance (ZSV). In this way, for all the proposed methods, a total of 100 features are extracted. In detail, 66 features belong to GLCM, 13 features belong to GLRLM, 13 features belong to GLSZM, 5 features belong to NGTDM, and 3 features belong to the GLOBAL method. Because of the difference in the range of feature variations, the features are normalized in the range of [0.1].

2.5. Classification Algorithms

In this study, various classification algorithms are utilized, including linear/non-linear Support Vector Machine (SVM) [21], Logistic Regression (LR) [22], Quadratic Discriminant Analysis (QDA) [23], Naïve Bayes (NB) [24], Random Forest (RF) [25], K-Nearest Neighbors (KNN) [26] and Multi-Layer Perceptron (MLP) [27]. SVM machine maps the training data space to a new one with the help of a non-linear mapping and then searches for a hypersurface that separates the samples of one class from the other class. According to the degree of complexity and entanglement of the data, common linear or non-linear polynomial mapping functions such as polynomial, Gaussian, and Radial Basis Functions (RBF) can be selected. The optimization process in SVM is to find the parameters of the hypersurface in the new feature space, which performs the data separation process in the best condition and with the least error. Briefly, the different steps of nonlinear support vector machines are to map the data range into the new feature space using kernel mapping functions and then partition the new feature space range by a hypersurface. This algorithm shows good performance in the classification of high-dimensional spaces. If the number of defined dimensions is more than the number of data and samples, the performance of this algorithm can be seen better. But on the other hand, this algorithm is not very suitable for very large data sets. A support vector machine is basically a binary separator. A multi-class problem can be solved by combining two-class support vector machines. The strategy used is one class versus the other classes to categorize each class. Then the outputs of the binary separators are combined and thus the multi-class problem is solved. The Logistic Regression method, as a statistical method, investigates the probability of each class. In fact, this algorithm performs as same as linear regression methods, but it has more complex cost functions and instead of a linear function between the coefficients, a sigmoid function is defined on them. This method, instead of relying on the values 0 and 1, results in possible values that are between zero and one. The effort of this algorithm is to identify a probability estimation between a series of independent variables and a dependent one. This method is one of the most widely used classification algorithms due to its simplicity of operation and easy interpretability. Another algorithm called QDA is a classifier with a polynomial decision boundary that maps a Gaussian function to the adaptive model of each class. This technique is generally used because it simultaneously implements the process of classification and dimensionality reduction of input data. The difference between this algorithm and LDA is the non-linearity of the decision boundary. Based on the conditions set for the training data, the discriminant function is statistically determined to deal with new data and relate them to different classes. The variables of this function are responsible for separation based on the variance in the class and between classes. According to Fisher's procedure [28], this algorithm increases inter-class variance compared to intra-class variance based on the mapping imposed on the desired data. Another classification algorithm called Naïve Bayes is also considered a probability-based classifier, which is based on Bayes' law. In this method, it is considered that each two extracted features of the studied samples for classification are independent of each other and every feature influences independently on assigning the output of the process which may not be satisfied in the real world. Therefore, by calculating the probability of the output with the conditions of the input features independently, it is possible to report a general probability of the conditions of the output for the possible status of the performance of the input features. The KNN algorithm, as a non-parametric approach, operates based on classification for a common feature of data in the neighborhood. This algorithm can be implemented simply and does not need to construct several models and adjust parameters or define additional assumptions. But it is highly dependent on determining the parameter K as

the selection of the number of samples in the neighborhood. Finally, decision tree-based methods are among the methods with simpler analysis and better response than other mathematical methods. In these algorithms, it is assumed that each selected category of features can provide an estimate of the final data class with the help of logical analysis of its changes compared to other features. For this classification method, different algorithms have been introduced. The random Forest algorithm utilizes a random set of features to build decision trees. The number of leaves, nodes, and branches is different, and the result is provided based on the majority voting method. The result of the RF algorithm is highly dependent on the number of extracted decision trees.

2.6. Statistical Analysis

For the evaluation of the results, some common analysis tools are used; for the multi-class system, accuracy is considered a measure of the performance of the used algorithm and can be used to report the performance of the entire network. In this work, the 10-fold cross-validation process is used for training and final testing. The number of training and testing data is chosen based on the percentage of 90% and 10% of the data, respectively. Also, due to the multi-class procedure, and as a result the need to check the efficiency of the algorithm for each class, the confusion matrix is calculated and reported for the average test data in each algorithm. In this matrix, for each class, correctly labeled samples and wrongly labeled samples are introduced. This matrix can comprehensively obtain all the evaluation criteria of different classes such as precision, recall, f1-score, sensitivity, and specificity.

3. Results

Table 2 presents the accuracy results obtained for different classification algorithms in desired pathological indices, including fibrosis, steatosis, lobular inflammation, and ballooning. According to the results, Gaussian SVM provides the highest accuracy and the lowest variance in grading the class in each of the four states of fibrosis, steatosis, ballooning, and lobular inflammation. This value can be seen in fibrosis with 6 classes, 84.30±0.66, in steatosis with 5 classes, 90.55±0.71, in lobular inflammation with 4 classes, 81.11±0.65, and in ballooning with 3 classes, 95.98±0.28. The results of fibrosis, steatosis, lobular inflammation, and ballooning are presented all together in Table 3 to check the algorithm's performance better. According to the better performance of the Gaussian SVM algorithm compared to other classification methods in Table 2, the results of the confusion matrix for this algorithm in different four pathological indices are as follows: a) fibrosis (in degrees of 0 to 4 with ignorable cases), b) steatosis (in degrees of 0 to 3 with ignorable cases), c) lobular inflammation (in degrees of 0 to 2 with ignorable cases), and d) ballooning (in degrees of 0 and 1 with ignorable cases) for test data in different classes are presented in Table 4.

4. Discussion

In this study, a combination of radiomics methods, including Global, GLCM, GLRLM, GLSZM, and NGTDM with the help of different machine learning methods, SVM, LR, QDA, KNN, RF, NB, and MLP is employed to classify the pathological images of the liver samples in four pathological indices, fibrosis (5 classes along with ignorable cases), Steatosis (4 classes with ignorable cases), Lobular inflammation (3 classes with ignorable cases), and ballooning (2 classes with ignorable cases), successfully. The highest accuracy and the lowest variance of the results were provided according to the Gaussian SVM method in the classification of fibrosis, steatosis, inflammation, and ballooning as 84.30±0.66, 90.55±0.71, 81.11±0.65, and 95.98±0.28, respectively.

In this study, SVM, RF, KNN, QDA, NB, MLP, and LR methods were chosen for classification. One of the reasons for choosing these algorithms is the multi-class nature of the studied data as well as the different mathematical and functional structures of the mentioned algorithms. The results show that the studied algorithms present a unique performance in four pathological indices. According to the results, it can be said that the Gaussian SVM method shows the highest accuracy and the lowest variance due to the consideration of a hyperplane with a non-linear distribution kernel (Table 2, Table 4). SVM machine maps the training data space to a new one with the help of a non-linear mapping and then searches for a hypersurface that separates the samples of one class from the other class. With suitable

Table 2. Comparison of accuracy obtained for different classification algorithms in pathological indices (a) fibrosis, (b) steatosis, (c) lobular inflammation, and (d) ballooning. In fibrosis, 6 classes, 5 classes in steatosis, 4 classes in lobular, and 3 classes in ballooning have been classified

a) Fibrosis

Model	Total ACC	Accuracy per class					
		0	1	2	3	4	ignore
LR	80.29±0.57	94.75	79.93	54.88	72.90	93.37	87.93
QDA	81.54±0.78	95.79	81.37	63.44	72.43	94.43	84.86
RF	84.01±0.59	94.88	81.18	63.29	79.78	94.59	92.18
KNN	83.96±0.75	93.40	82.47	62.34	77.60	95.71	94.14
NB	68.10±0.92	88.11	67.72	61.70	27.73	89.06	78.39
LSVM	82.10±0.86	96.02	81.57	53.57	80.59	92.88	89.65
GSVM	**84.30±0.66**	95.41	82.79	58.54	84.01	94.46	92.11
MLP	68.67±1.06	89.95	63.98	40.27	57.01	89.34	78.59

b) steatosis

Model	Total ACC	Accuracy per class				
		0	1	2	3	ignore
LR	90.39±0.74	92.96	86.94	83.27	93.54	96.00
QDA	90.46±0.78	94.08	88.96	80.33	96.37	94.81
RF	89.77±0.60	93.19	84.09	82.68	93.86	95.13
KNN	88.46±1.06	92.04	83.72	82.38	92.35	92.25
NB	82.06±1.47	89.90	86.11	64.24	93.93	83.15
LSVM	90.51±0.77	94.20	87.10	82.36	94.93	96.63
GSVM	**90.55±0.71**	94.35	85.31	83.44	94.52	95.71
MLP	81.17±1.29	89.74	65.46	71.59	93.29	83.86

c) Lobular inflammation

Model	Total ACC	Accuracy per class			
		0	1	2	ignore
LR	79.44±0.66	90.32	64.14	74.38	89.67
QDA	80.07±0.97	93.17	71.21	70.95	87.25
RF	80.19±0.72	90.04	64.53	76.90	90.85
KNN	78.63±0.96	87.75	60.98	73.15	93.17
NB	64.93±0.47	85.24	66.50	26.65	86.40
LSVM	79.74±0.58	90.93	64.53	74.20	90.12
GSVM	**81.11±0.65**	91.52	66.61	76.33	90.63
MLP	62.78±1.90	85.78	58.81	28.15	86.72

d) ballooning

Model	Total ACC	Accuracy per class		
		0	1	ignore
LR	94.44±0.31	94.86	91.95	96.67
QDA	94.39±0.47	94.82	94.24	94.28
RF	94.28±0.31	93.88	92.46	96.61
KNN	94.65±0.34	92.79	93.73	97.45
NB	85.77±0.54	79.27	89.72	87.94
LSVM	94.79±0.37	96.10	91.77	96.72
GSVM	**95.98±0.28**	96.23	94.68	97.10
MLP	86.83±2.39	88.23	83.24	92.15

Table 3. Comparison of the algorithms' performance for the fibrosis, steatosis, lobular inflammation, and ballooning

Model	Fibrosis	Steatosis	Inflammation	Ballooning
LR	80.29	90.39	79.44	94.44
QDA	81.54	90.46	80.07	94.39
RF	84.01	89.77	80.19	94.28
KNN	83.96	88.46	78.63	94.65
NB	68.10	82.06	64.93	85.77
LSVM	82.10	90.51	79.74	94.79
GSVM	**84.30**	**90.55**	**81.11**	**95.98**
MLP	68.67	81.17	62.78	86.83

Table 4. The results of the Confusion Matrix for Gaussian SVM algorithm in different four pathological indices: a) fibrosis (in grades 0 to 4 and ignorable cases), b) Steatosis (in grades 0 to 3 with ignorable cases), c) lobular inflammation (in grades 0 to 2 with ignorable cases), and d) ballooning (in grades 0 and 1 with ignorable cases) for test data

a

True Class \ Predicted Class	L0	L1	L2	L3	L4	Ignore
L0	219	4	1	3	0	3
L1	2	214	24	18	0	1
L2	3	31	142	63	0	3
L3	7	7	22	192	2	1
L4	0	0	5	7	209	0
Ignore	5	2	9	3	0	225

b

True Class \ Predicted Class	L0	L1	L2	L3	Ignore
L0	111	6	0	0	1
L1	5	118	15	0	1
L2	0	8	224	36	0
L3	0	0	14	244	0
Ignore	3	2	2	3	218

c

True Class \ Predicted Class	L0	L1	L2	Ignore
L0	506	23	20	4
L1	36	382	150	6
L2	15	132	498	7
Ignore	8	25	23	541

d

True Class \ Predicted Class	L0	L1	Ignore
L0	1149	31	24
L1	59	1226	10
Ignore	19	17	1182

nonlinear mapping, these data sets are separated well by a hyperplane. The advantage of using the new feature space is that non-separable classes can be separated by choosing a suitable kernel function. This algorithm shows good performance in the classification of high-dimensional spaces. Finally, after the Gaussian SVM, the RF algorithm shows the highest accuracy compared to the other algorithms. The RF algorithm utilizes a random set of features to build different decision trees and the result is provided based on the majority voting method of these extracted decision trees. So, the RF tries to consider different strategies to assign labels based on various arrangements of the features.

One of the frequent problems in multi-class diagnostic processes is the separation of intermediate stages in the images. In these stages, accurately determining the class type is so difficult due to the homogeneity of pixel variations in the pathological images. If we look carefully at the results of the Confusion Matrix in Table 4, we can see that it is not possible to fully separate between the middle classes, for example, between classes 2 and 3 in lobular inflammation, 3 and 4 in steatosis, and 1, 2, and 3 in fibrosis, and the number of incorrect classifications is considerable in these categories. Therefore, the Gaussian SVM classification algorithm, like other classification methods, has

problems in the middle classes and shows its greatest error in classification in these classes.

Table 5 provides a comparison of the results of similar studies with the present results. The results show that the proposed algorithms have been able to achieve acceptable results with the help of the extracted texture features and the performance of the Gaussian SVM method in the main indices. Our strength of the proposed method is the interpretability of the algorithm along with its simplicity and low time-consuming. The complexity of our method is less than Convolutional Neural Networks (CNN) and the requirement for a special processing system is eliminated. These CNN algorithms require massive data sets to train the models properly. Their interpretability is an alternative challenge that should be considered in choosing the desired algorithms. Therefore, our algorithm can be considered a simpler solution for proper classification in various pathological indices, and in the future, by extracting and selecting more effective features by feature selection methods, the results can be improved.

Table 5. Comparison of similar studies with the present results

Ref	No. of candidates	Modality	ML algorithm	Steatosis	Fibrosis	Inflammation	Ballooning
Jana [11]	30	Pathological images + CT scan	ResNet-18	95.95	89.84	74.46	77.19
Arjmand [12]	720	Pathological images	CNN + SoftMax	90	-	-	-
Yang [14]	36	Pathological images	CNN	88.34	-	-	-
Heinemann [17]	258	Pathological images	InceptionResNet-v3 + SoftMax	94.5	86.3	86	93.1
Teramoto [1]	79	Pathological images	Logistic Regression	-	-	-	97.8
Proposed algorithm	258	Pathological images	{GLCM, GLRLM, GLSZM, …} + {LR, SVM, RF, …}	90.55	84.30	81.11	95.98

5. Conclusion

In this study, a combination of the features obtained from Global, GLCM, GLRLM, GLSZM, and NGTDM methods and the Gaussian SVM classification algorithm is considered successful to classify different pathological images of liver tissue in indices of fibrosis, steatosis, lobular inflammation, and ballooning. Therefore, this fully automatic framework based on advanced machine learning algorithms from pathological images can be very useful in clinical procedures and can be used as an assistant or a substitute for pathologists' diagnoses with an acceptable level of accuracy in various examinations.

Acknowledgments

This research is financially supported by "Research Department of School of Medicine Shahid Beheshti University of Medical Sciences" [Grant No 21340].

References

1- Zobair Younossi *et al.*, "Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis." *Hepatology*, https://doi.org/10.1002/hep.30251 Vol. 69 (No. 6), pp. 2672-82, 2019/06/01 (2019).

2- K. Dyson Jessica, M. Anstee Quentin, and McPherson Stuart, "Non-alcoholic fatty liver disease: a practical approach to diagnosis and staging." *Frontline Gastroenterology*, Vol. 5 (No. 3), p. 211, (2014).

3- David E. Kleiner *et al.*, "Design and validation of a histological scoring system for nonalcoholic fatty liver disease." *Hepatology*, https://doi.org/10.1002/hep.20701 Vol. 41 (No. 6), pp. 1313-21, 2005/06/01 (2005).

4- Matthew M. Yeh and Elizabeth M. Brunt, "Pathological Features of Fatty Liver Disease." *Gastroenterology*, Vol. 147 (No. 4), pp. 754-64, 2014/10/01/ (2014).

5- Takuya Kuwashiro *et al.*, "Discordant pathological diagnosis of non-alcoholic fatty liver disease: A prospective multicenter study." *JGH Open*, https://doi.org/10.1002/jgh3.12289 Vol. 4 (No. 3), pp. 497-502, 2020/06/01 (2020).

6- Ramón Bataller and David A. Brenner, "Liver fibrosis." *The Journal of Clinical Investigation*, Vol. 115 (No. 2), pp. 209-18, 02/01/ (2005).

7- Yu Sub Sung, Bumwoo Park, Hyo Jung Park, and Seung Soo Lee, "Radiomics and deep learning in liver diseases." *Journal of Gastroenterology and Hepatology*, https://doi.org/10.1111/jgh.15414 Vol. 36 (No. 3), pp. 561-68, 2021/03/01 (2021).

8- Roi Anteby *et al.*, "Deep learning for noninvasive liver fibrosis classification: A systematic review." *Liver International*, https://doi.org/10.1111/liv.14966 Vol. 41 (No. 10), pp. 2269-78, 2021/10/01 (2021).

9- Amaro Taylor-Weiner *et al.*, "A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH." *Hepatology*, https://doi.org/10.1002/hep.31750 Vol. 74 (No. 1), pp. 133-47, 2021/07/01 (2021).

10- Mousumi Roy *et al.*, "Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies." *Laboratory Investigation*, Vol. 100 (No. 10), pp. 1367-83, 2020/10/01 (2020).

11- A. Jana, H. Qu, P. Rattan, C. D. Minacapelli, V. Rustgi, and D. Metaxas, "Deep Learning based NAS Score and Fibrosis Stage Prediction from CT and Pathology Data." in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, (2020), pp. 981-86.

12- A. Arjmand *et al.*, "Deep Learning in Liver Biopsies using Convolutional Neural Networks." in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, (2019), pp. 496-99.

13- Roberta Forlano *et al.*, "High-Throughput, Machine Learning–Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease." *Clinical Gastroenterology and Hepatology*, Vol. 18 (No. 9), pp. 2081-90.e9, 2020/08/01/ (2020).

14- Fan Yang *et al.*, "Quantification of hepatic steatosis in histologic images by deep learning method." *Journal of X-Ray Science and Technology*, Vol. 27pp. 1033-45, (2019).

15- Hui Qu *et al.*, "Training of computational algorithms to predict NAFLD activity score and fibrosis stage from liver histopathology slides." *Computer Methods and Programs in Biomedicine*, Vol. 207p. 106153, 2021/08/01/ (2021).

16- Yumeng x., "A machine learning-based approach for quantitative and automated non-alcoholic fatty liver disease (NAFLD)/non-alcoholic steatohepatitis (NASH) assessment using pathological stained slides." *National University of Singapore*, (2019).

17- Fabian Heinemann, Gerald Birk, and Birgit Stierstorfer, "Deep learning enables pathologist-like scoring of NASH models." *Scientific Reports*, Vol. 9 (No. 1), p. 18454, 2019/12/05 (2019).

18- Fabian Heinemann *et al.*, "Deep learning-based quantification of NAFLD/NASH progression in human liver biopsies." *Scientific Reports*, Vol. 12 (No. 1), p. 19236, 2022/11/10 (2022).

19- https://www.zeiss.com/microscopy/int/products/microscope-software/zen.html.

20- Antreas Antoniou, Amos Storkey, and Harrison Edwards, "Data augmentation generative adversarial networks." *arXiv preprint arXiv:1711.04340*, (2017).

21- Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends." *Neurocomputing*, Vol. 408pp. 189-215, (2020).

22- Kanish Shah, Henil Patel, Devanshi Sanghvi, and Manan Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification." *Augmented Human Research*, Vol. 5pp. 1-16, (2020).

23- Benyamin Ghojogh and Mark Crowley, "Linear and quadratic discriminant analysis: Tutorial." *arXiv preprint arXiv:1906.02590*, (2019).

24- Feng-Jen Yang, "An implementation of naive bayes classifier." (2018): *IEEE*, pp. 301-06.

25- Anjaneyulu Babu Shaik and Sujatha Srinivasan, "A brief survey on random forest ensembles in classification model." (2019): *Springer*, pp. 253-60.

26- Antonio Mucherino, Petraq J. Papajorgji, Panos M. Pardalos, Antonio Mucherino, Petraq J. Papajorgji, and Panos M. Pardalos, "K-nearest neighbor classification." *Data mining in agriculture*, pp. 83-106, (2009).

27- Arti Rana, Arvind Singh Rawat, Anchit Bijalwan, and Himanshu Bahuguna, "Application of multi layer (perceptron) artificial neural network in the diagnosis system: a systematic review." (2018): *IEEE*, pp. 1-6.

28- Yunlei Li, Lodewyk F. A. Wessels, Dick de Ridder, and Marcel J. T. Reinders, "Classification in the presence of class noise using a probabilistic kernel fisher method." *Pattern Recognition*, Vol. 40 (No. 12), pp. 3349-57, (2007).

29- Takashi Teramoto, Toshiya Shinohara, and Akihiro Takiyama, "Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology." *Computer Methods and Programs in Biomedicine*, Vol. 195p. 105614, (2020).