

Article

Efficiency of Machine Learning Algorithms for the Determination of Macrovesicular Steatosis in Frozen Sections Stained with Sudan to Evaluate the Quality of the Graft in Liver Transplantation

Fernando Pérez-Sanz ^{1,*}, Miriam Riquelme-Pérez ², Enrique Martínez-Barba ³, Jesús de la Peña-Moral ³, Alejandro Salazar Nicolás ³, Marina Carpes-Ruiz ⁴, Angel Esteban-Gil ¹, María Del Carmen Legaz-García ¹, María Antonia Parreño-González ¹, Pablo Ramírez ^{5,*} and Carlos M. Martínez ⁴



Citation: Pérez-Sanz, F.; Riquelme-Pérez, M.; Martínez-Barba, E.; de la Peña-Moral, J.; Salazar Nicolás, A.; Carpes-Ruiz, M.; Esteban-Gil, A.; Legaz-García, M.D.C.; Parreño-González, M.A.; Ramírez, P.; et al. Efficiency of Machine Learning Algorithms for the Determination of Macrovesicular Steatosis in Frozen Sections Stained with Sudan to Evaluate the Quality of the Graft in Liver Transplantation. *Sensors* **2021**, *21*, 1993. <https://doi.org/10.3390/s21061993>

Academic Editor: Mukesh Prasad

Received: 27 January 2021

Accepted: 10 March 2021

Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¹ Biomedical Informatics & Bioinformatics Service, Institute for Biomedical Research of Murcia (IMIB), 30120 Murcia, Spain; angel.esteban@ffis.es (A.E.-G.); mcarmen.legaz@ffis.es (M.D.C.L.-G.); maria.parreno@ffis.es (M.A.P.-G.)
- ² CNRS-CEA, University Paris-Saclay, MIRCen, 92265 Paris, France; miriam.riquelme-perez@cea.fr
- ³ Pathology Service, University Clinical Hospital Virgen de la Arrixaca-Biomedical Research Institute of Murcia (IMIB), 30120 Murcia, Spain; enrique.martinez4@carm.es (E.M.-B.); jesus@biopathocid.com (J.d.I.P.-M.); alejandro.salazar@um.es (A.S.N.)
- ⁴ Experimental Pathology Service, Institute for Biomedical Research of Murcia (IMIB), 30120 Murcia, Spain; marina.carpes@imib.es (M.C.-R.); carlos.martinez@imib.es (C.M.M.)
- ⁵ General and Digestive Surgery Service, University Clinical Hospital Virgen de la Arrixaca-Biomedical Research Institute of Murcia (IMIB), 30120 Murcia, Spain
- * Correspondence: fernando.perez@ffis.es (F.P.-S.); pablo.ramirez@carm.es (P.R.)
- † Current address: Biomedical and Bioinformatics Service, Institute for Biomedical Research of Murcia (IMIB), Crta. Buenavista s/n, 30120 Murcia, Spain.
- ‡ These authors contributed equally to this work.

Abstract: Liver transplantation is the only curative treatment option in patients diagnosed with end-stage liver disease. The low availability of organs demands an accurate selection procedure based on histological analysis, in order to evaluate the allograft. This assessment, traditionally carried out by a pathologist, is not exempt from subjectivity. In this sense, new tools based on machine learning and artificial vision are continuously being developed for the analysis of medical images of different typologies. Accordingly, in this work, we develop a computer vision-based application for the fast and automatic objective quantification of macrovesicular steatosis in histopathological liver section slides stained with Sudan stain. For this purpose, digital microscopy images were used to obtain thousands of feature vectors based on the RGB and CIE L*a*b* pixel values. These vectors, under a supervised process, were labelled as fat vacuole or non-fat vacuole, and a set of classifiers based on different algorithms were trained, accordingly. The results obtained showed an overall high accuracy for all classifiers (>0.99) with a sensitivity between 0.844 and 1, together with a specificity >0.99. In relation to their speed when classifying images, KNN and Naïve Bayes were substantially faster than other classification algorithms. Sudan stain is a convenient technique for evaluating ME in pre-transplant liver biopsies, providing reliable contrast and facilitating fast and accurate quantification through the machine learning algorithms tested.

Keywords: liver transplantation; sudan stain; machine learning; computer vision; macrovesicular steatosis

1. Introduction

Liver transplantation is the unique curative option for those patients with end-stage liver disease and acute liver failure. The progressive demand for transplants and the limited number of available organs have led to modification of the scoring systems used

to assess post-transplant complications, in order to include livers excluded according to the older systems—such as organs from cardiac death (CDC), and HIV- and/or hepatitis C-infected patients—into the donor pool [1]. However, this fact also implies that the analytic procedures to evaluate the suitability of the grafts must also be more accurate. Therefore, histological analysis to assess the quality of the allografts is crucial to prevent graft dysfunction or secondary rejection.

As there is a strong relationship between the ischemia time—defined as the time from graft clamping to pre-implant reperfusion—and the risk of graft failure [2], this analysis must be assessed as soon as possible, in order to prevent severe liver damage. Thus, the use of hematoxylin and eosin (H & E) stained sections from frozen representative graft samples, instead of routine-processed ones, is a time effective alternative for this determination, as it allows for a rapid histological examination [3]. During histological analysis, several parameters should be assessed. One such parameter is the presence of big intracytoplasmic lipid droplets—macrovesicular steatosis (ME)—a feature which has shown predictive value for graft dysfunction [4]. Thus, livers with <30% ME are generally considered acceptable for transplantation [5,6], although this value varies by institution and can be increased up to 60% [7]. Nevertheless, this criteria is based on subjective evaluation by an experienced pathologist, which is strongly observation-dependent, non-reproducible, and challenging, even in experienced hands [8,9]. Additionally, the main limitation of the use of H&E frozen-stained sections is the risk of the underestimation of ME, due to the presence of artifacts (e.g., water droplets) during the sampling procedure [3,10]. Thus, the development of alternative technical and analytic procedures for staining and examining representative frozen graft sections, which allow for the establishment of an objective ME value in the shortest possible time, should be key to ensuring the viability of the transplant.

The development of image analysis tools based on machine learning algorithms for histopathologic analysis is a extremely helpful application of computational biology, helping pathologists to establish exact and accurate diagnoses. In this sense, several image analysis algorithms to determine the degree of ME of the graft have been developed, in terms of the determination of cross-sectional surface area of lipid droplets (LD) [11], the determination of this area using a pre-determined cut-off LD size [12,13], and the use of LD-induced nuclear dislocation of hepatocytes as methods to improve the algorithms [14], or even the use of a four-stage approach, including k-means clustering and image manipulation algorithms to detect fat areas [14]. Nevertheless, the main limitation of all of these applications is that their analytic procedures are based on H&E stained sections, which may induce an underestimation of the value [3,15]. Although several deep learning segmentation algorithms [10,16,17] and supervised machine learning procedures [18] have been proposed to improve the accuracy of such algorithms in classifying intracellular fat vacuoles—identified as white spaces in H&E sections—the use of these applications is still not fully standardized. Other solutions have been proposed, such as the assessment of liver steatosis by liver texture analysis, a macroscopic determination which uses machine learning to speed up and automate the process [19]; a non-microscopic parameter which is still under validation. Thus, the use of alternative and specific staining procedures, which allow for the development of simpler, quicker, and specific image analysis algorithms to determine ME in biopsies, is crucial in establishing an objective diagnosis in the shortest possible time.

Consequently, the use of specific fat-staining procedures is the first stage to overcome the limitations of H&E staining. Thus, the Sudan staining procedure may provide a good alternative, as it is a fat-specific, quick, and easy stain procedure which is performed only on frozen sections, showing higher sensitivity for the detection of steatosis, compared to H & E [15]. Additionally, and due to the limited time available to perform this analysis prior to the surgery, it is also necessary to optimize the analytic speed, the accuracy, and the computational cost of the specific image analysis technique. Thus, the time required to obtain the image set, the quality and the size of the images, and the time needed to obtain the final result are variables which must also be optimized. Therefore, the aim of this

report is to develop an image analysis application, based on machine learning algorithms, to automatically determine the percentage of ME in a representative sample from a donor liver using Sudan stained frozen sections, which may allow for optimized accuracy with the lowest image quality possible and, thus, lower computational costs (by using the minimum system requirements possible, even allowing for the use of an on-line application for the analysis), through pre-existing and validated machine/deep learning algorithms.

2. Results

2.1. Train and Test Model

The application was able to easily differentiate specific fat staining from artifacts related to the staining procedure. The average time that each algorithm takes to be trained is a function of the number of pixels involved in the training, which ranged from 1000 to 100,000 pixels for each model (Figure 1). Our results showed that using images of Sudan stained pre-transplant human donor liver sections, the training time using Keras increased drastically, from 0.64 s with 1000 pixels to 38 s with 100,000 pixels (Table 1). On the other hand, Naïve Bayes and KNN were the fastest algorithms in training stage, with the longest times of 0.011 and 0.006 s with 100,000 pixels, respectively.

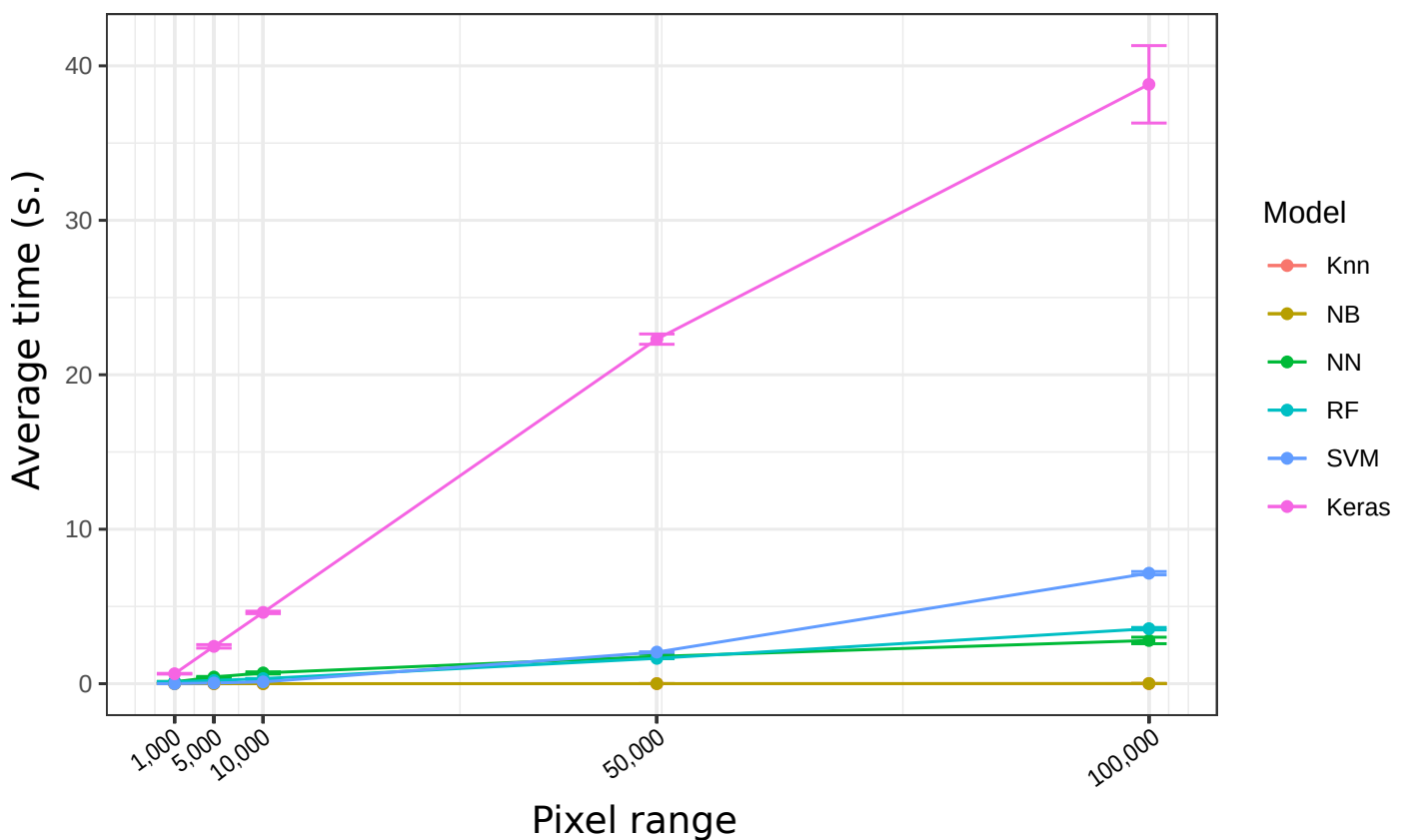


Figure 1. Average training time for each algorithm (in s), according with the proposed range of pixels selected.

Table 1. Average training time and AUC for each algorithm under different numbers of pixels.

Model	Pixels	Average Time	Average AUC	SE Time	SE AUC
KNN	1000	0.000	1.000	0.000	0.000
NB	1000	0.001	1.000	0.000	0.000
NN	1000	0.117	1.000	0.032	0.000
RF	1000	0.118	1.000	0.001	0.000
SVM	1000	0.004	1.000	0.000	0.000
Keras	1000	0.645	0.984	0.022	0.012

Table 1. Cont.

Model	Pixels	Average Time	Average AUC	SE Time	SE AUC
KNN	5000	0.000	0.997	0.000	0.000
NB	5000	0.001	0.998	0.000	0.000
NN	5000	0.429	1.000	0.040	0.000
RF	5000	0.207	0.999	0.002	0.000
SVM	5000	0.045	1.000	0.000	0.000
Keras	5000	2.417	0.998	0.113	0.000
KNN	10,000	0.001	0.998	0.000	0.000
NB	10,000	0.001	0.999	0.000	0.000
NN	10,000	0.697	1.000	0.069	0.000
RF	10,000	0.329	0.998	0.002	0.000
SVM	10,000	0.115	1.000	0.001	0.000
Keras	10,000	4.616	0.999	0.080	0.000
KNN	50,000	0.003	0.997	0.000	0.000
NB	50,000	0.005	0.998	0.000	0.000
NN	50,000	1.763	0.999	0.141	0.000
RF	50,000	1.642	0.999	0.011	0.000
SVM	50,000	2.046	0.999	0.011	0.000
Keras	50,000	22.307	0.999	0.333	0.000
KNN	100,000	0.006	0.997	0.000	0.000
NB	100,000	0.011	0.997	0.001	0.000
NN	100,000	2.799	0.999	0.210	0.000
RF	100,000	3.562	0.999	0.079	0.000
SVM	100,000	7.153	0.999	0.107	0.000
Keras	100,000	38.802	0.999	2.507	0.000

The trained classifiers were tested with a test data set, comprised of 30% of the sample corresponding to each set of pixels. Overall, the performance achieved by the classifiers with different sets of pixels was close to 1. Thus, the analysis of the area under the curve (AUC) obtained from the different ROC curves was also close to 1 (Figures 2a–e). Only Keras with 0.98 had an AUC below 0.99 for the 1000-pixel proposed model (Table 1, Figure 2a).

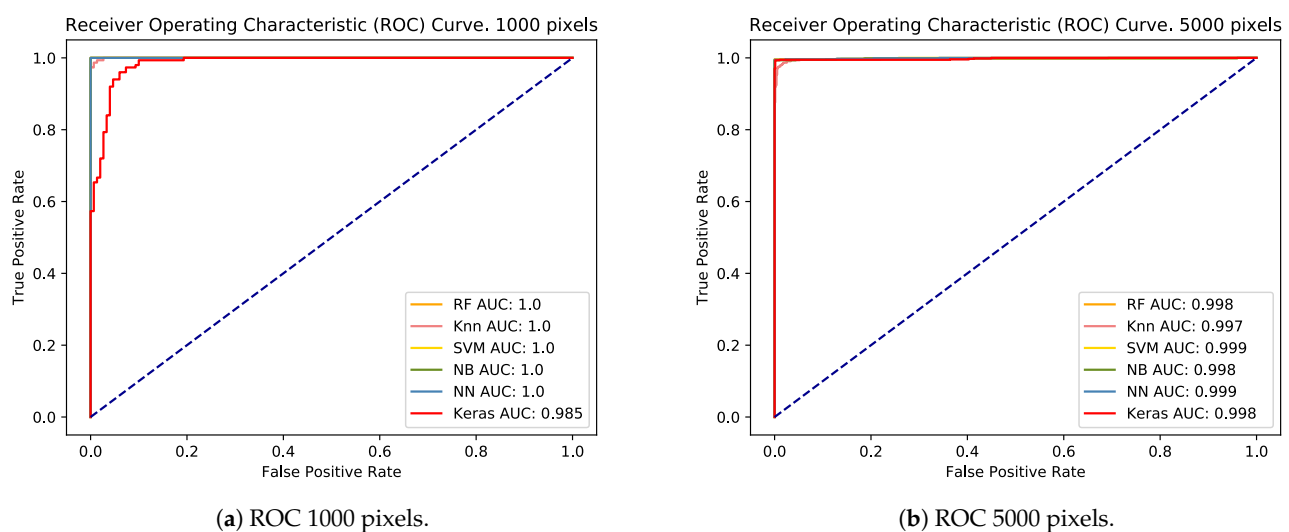


Figure 2. Cont.

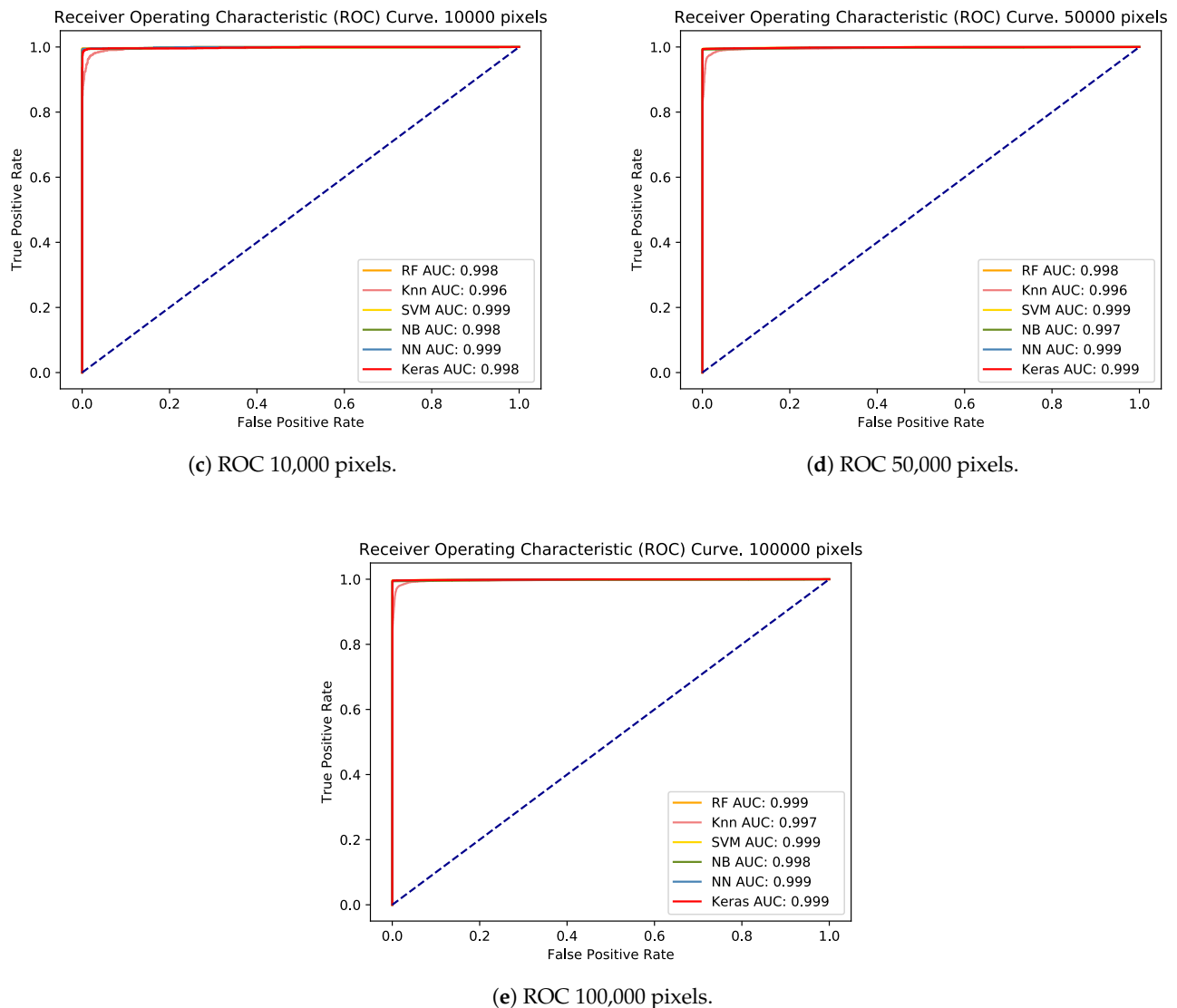


Figure 2. ROC curves with the AUCs of all classifiers for each training/testing data set from 1000 (a) to 100,000 (e) pixels.

2.2. Classification Time

The time taken to classify one image differed widely, depending on the classifier and the image size (from 1.5 to 6.1 MB; Table 2 and Figure 3). Regardless of the image size, KNN and NB were the fastest algorithms, being slightly affected by both the number of threads and the image size. On the other hand, SVM and RF were the slowest algorithms, which were strongly affected by the size of the image and the number of processing threads. As Keras handles the number of threads independently, due to its implementation, it always involves all threads of the processor. Thus, the average classification time with this algorithm was always the same (Figure 3).

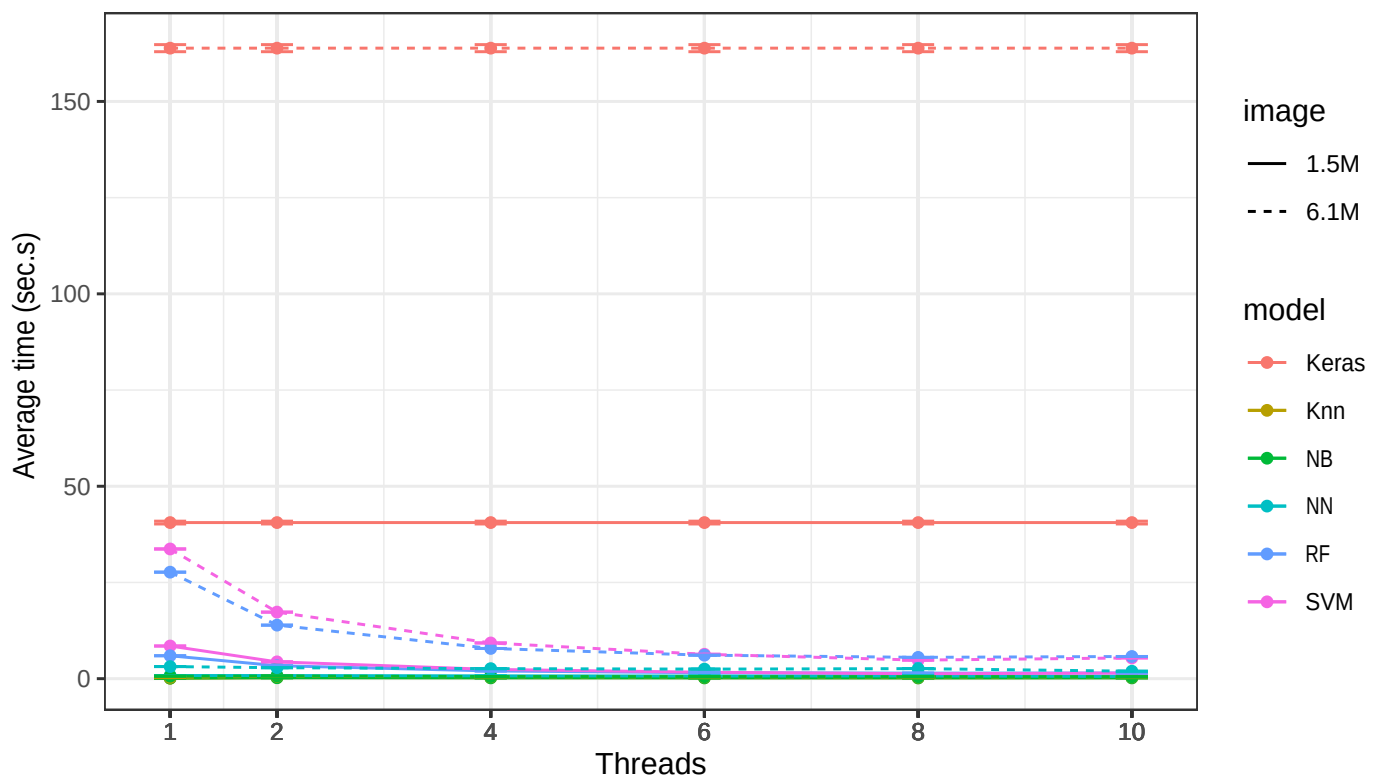


Figure 3. Classification time (in s) of each model used.

Table 2. Average classification time (in s), based on number of threads (from 1 to 10) and image size (from 1.5 to 6.1 MB).

Model	Image	1	2	4	6	8	10
KNN	1.5 M	0.09	0.28	0.26	0.24	0.22	0.26
SVM	1.5 M	8.48	4.37	2.41	1.67	1.29	1.46
RF	1.5 M	5.96	3.37	2.01	1.55	1.44	1.47
NB	1.5 M	0.15	0.25	0.22	0.21	0.20	0.21
NN	1.5 M	0.82	0.82	0.73	0.69	0.72	0.57
Keras	1.5 M	40.56	40.56	40.56	40.56	40.56	40.56
KNN	6.1 M	0.32	0.61	0.49	0.45	0.42	0.46
SVM	6.1 M	33.70	17.30	9.28	6.34	4.82	5.39
RF	6.1 M	27.68	13.91	7.85	6.10	5.54	5.77
NB	6.1 M	0.66	0.69	0.55	0.52	0.51	0.51
NN	6.1 M	3.14	2.85	2.61	2.49	2.62	1.94
Keras	6.1 M	163.84	163.84	163.84	163.84	163.84	163.84

2.3. Classification Validation

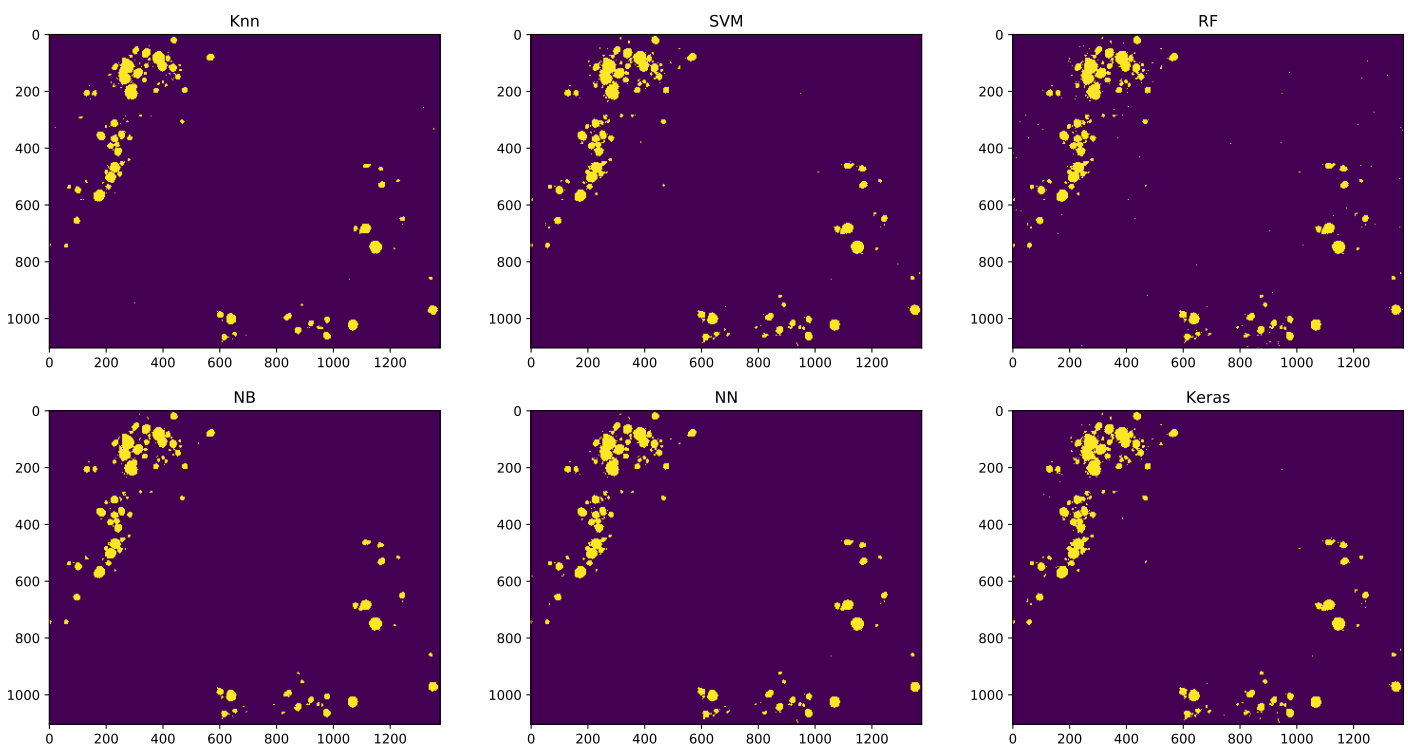
Images classified by every model were compared with the same image manually classified by an expert pathologist (Figure 4).

Based on this comparison between images, we have calculated the accuracy (ratio of well-classified pixels to total pixels; Equation (2)), sensitivity (ratio of detected fatty vacuole pixels to total fatty vacuole pixels; Equation (3)), and specificity (ratio of non-fatty vacuole pixels detected to total non-fatty pixels; Equation (4)).

In all cases (Table 3), the accuracy, sensitivity, and specificity values were above 0.95; except for KNN, whose sensitivity was 0.844. These values were consistent with those obtained from evaluating classifiers with the train/test data sets, as shown by the ROC curves.

Table 3. Metrics comparing automatic and manual classification.

Metric	KNN	SVM	RF	NB	NN	Keras
Accuracy	0.996	0.996	0.996	0.997	0.997	0.995
Sensitivity	0.844	0.962	0.956	0.910	0.963	0.972
Specificity	0.999	0.997	0.997	0.999	0.998	0.996
Precision	0.961	0.897	0.894	0.969	0.906	0.856

**Figure 4.** Results of image classification for each classifier.

3. Discussion

Our goal was to develop an application which is able to establish an objective and reliable value of macrovesicular steatosis from representative sections of pre-transplant liver donor biopsies stained with Sudan—a fat-specific staining procedure—with minimum requirements, in terms of image quality and processing time. For this purpose, we tested several classification machine learning algorithms, in order to determine which algorithm is the most suitable for application. To the best of our knowledge, this is the first report in which several machine learning algorithms were tested for the automatic analysis of fat-specific dye stained sections for biomedical purposes. Moreover, we developed a graphical user interface (GUI) implementing the algorithms discussed in this work. It also allows for the training of new models, based on the same algorithms. The development framework used was *Shiny*—a web development framework based on R—which allows for near-native integration of all Python code necessary for generating the models and analyzing the images. The simple and intuitive design makes it easy for the end-user to quickly quantify steatosis.

Although the number of potential donors for liver transplant has increased, the number of canceled transplantations due to a high grade of ME have also risen [20]. As this parameter still must undergo a subjective evaluation, the possibility of an error of criteria cannot be excluded; even in the case of analysis by an expert pathologist. Liver transplantation is an extremely complex surgery, the success of which depends on the time consumed between organ extraction from the donor and its reperfusion into the patient. Thus, the intraoperative histopathologic evaluation—which usually involves sampling, sectioning,

staining, examination, and diagnosis—must be assessed in less than 30–45 min [21]. As the fastest fixation and paraffin embedding procedures usually require 3–4 h, the use of frozen samples is mandatory in this case. H & E is usually the standard procedure for general evaluation, which is easy and quick to perform and usually provides a good contrast to evaluate many parameters used to establish the quality of the graft for transplant. Nevertheless, this procedure does not stain fat, and the possibility of overestimating ME due to artifacts produced during the processing of the frozen biopsies (e.g., water droplets, holes, and so on) may be not discarded [3,10]. Taking this into account, coupled with the fact that ME determination is strongly observation-dependent, we find that the risk of error of judgement can increase significantly, with severe consequences, regardless of the final decision.

Thus, improvement of the staining procedure and the accuracy of the steatosis determination, by transforming an estimated determination into a quantitative one, in the shortest time possible may allow for a drastically diminished possibility of error, an increase in the number of viable organs, and the establishment of more accurate outcomes, in terms of viability of the graft.

As the use of frozen sections to the immediate diagnosis is mandatory, our first goal was to use an alternative staining procedure, which may replace the H & E stain and allow for fat to be stained specifically. To this end, we decided to use the Sudan stain, as it can be performed on frozen sections, is a fast and easy stain procedure, and is fat-specific, making it chromatically easy to differentiate fat from non-fat structures. A possible disadvantage of this stain procedure is that the value of steatosis can be overestimated by direct examination, especially when the analysis depends on inexperienced pathologists. We did not observe significant variations in ME values during the validation process, probably due to the use of two experienced pathologists specialists in liver transplantation.

Once we had solved this problem, our next issue was to determine which is the best machine learning algorithm for use in the development of our analytic tool. As all reported image analysis tools have been based on the analysis of H & E stained sections [11–14], these algorithms are focused on the measurement of numerous parameters, which try to differentiate structures (i.e., fat vacuoles vs. non-fat vacuoles and unspecific structures) with similar shape and color (i.e., round unstained/white structures). As there have been no previous reports considering the use of Sudan stain for the automatic determination of ME, we decided to use six of the most-used algorithms for image analysis [22], in order to determine which is the best option—in terms of efficacy and time—for use in a new and specific image system based on Sudan-stained section analysis. Additionally, we took into account the time used for the analysis, as this parameter should not be extended, in order to assure the efficiency of the procedure. Thus, the use of high-resolution scanned images may be not useful in this particular case, as the time required to obtain and process such images (which are near 1 GB in size each) may be not applicable to study one parameter, which must be objectively determined in 5–10 min at maximum. For this reason, our tool does not currently use whole slide images, although we are considering their use as future work, provided that their processing time can be optimised. Therefore, our goal was to develop an image analysis application with the best machine learning algorithm, which is able to establish an objective value of ME using the lowest image resolution possible, in order to optimize either the processing time and/or the requirements of the system employed in the analysis (potentially even providing the possibility of performing through an on-line web application). The evaluated algorithms showed high performance, in terms of image classification. In the training/testing phase, the AUC obtained in all cases was significantly high (>0.98), which was virtually unaffected by the number of pixels used. Only the 1000 pixel data set decreased the AUC of the Keras algorithm. In the trials carried out, it was shown that the AUC for 1000 pixels was more affected by the pixels selected in the random sampling than by the number of pixels used.

Concerning the time spent by each classifier to be trained, it is worth mentioning the outstanding difference between Keras and the remaining algorithms. Furthermore,

for each classifier, the time increase was more significant from 50,000 pixels onwards; except with KNN and NB, whose times were barely affected. Thus, between 10,000 and 50,000 pixels, a compromise can be achieved between time spent in training and robustness in the random sampling of pixels.

In the classification step of a real image, KNN and NB were the fastest, regardless of the number of threads. Even in the case of an image with 4 times more pixels than another, the duration was shorter than 1 s. On the other hand, RF and SVM were significantly influenced by both the size of the image and the number of threads involved in the classification. Nevertheless, from six threads on, there was no significant reduction in the time spent classifying the image on the equipment used; thus, it may be unnecessary to invest more computational resources, when the performance is not going to be enhanced substantially.

The global result, when comparing the manually classified and the same automatically classified image with the different classifiers, yielded good results overall. The ratio of positives over the total number of images (i.e., accuracy) was close to 1 in every case. With regard to the sensitivity of Keras, it was the most accurate, with 100% success; while KNN, with 0.844 accuracy, was the most mistaken. The specificity, similar to the accuracy, remained very high for all the classifiers. The main limitation regarding the use of these algorithms was observed in those cases with an extremely high infiltration of fat (70–80%). In those cases, the sensibility of the algorithms to pixel selection was increased, possibly due to the fat infiltration observed within portal and stromal areas. In such cases, alternatives like the use of morphological segmentation algorithms would be helpful to establish accurate values of ME. Another limitation was based on the fact that we did not compare the accuracy of these classification algorithms with other lipid staining methods, such as the oil red procedure; however, our results indicate that the use of specific fat staining procedures, such as Sudan, may be a good choice for the automatic determination of ME in pre-transplant liver biopsies, using minimal requirements with optimal results for those cases with low–average amounts of fat infiltration. Such cases are those in which the pathologist may experience problems in establishing an accurate value of ME.

We conclude, based on these results, that Sudan stain is a suitable and value tool for the evaluation of ME in pre-transplant liver biopsies, as it is suitable for frozen sections, quick, fat-specific, and offers good contrast, which can allow for easy differentiation of fat vacuoles from non-fat vacuoles and unspecific structures; while H & E stain can be used for the evaluation of all other parameters, such as inflammation, infection, necrosis, tumors, and pigment deposits. We propose the introduction of this stain as the technique to use in the evaluation of ME in such biopsies. As our goal was to develop an automatic analytic system which may determine the amount of ME in a stained section, this also may save valuable time for the pathologist, in terms of evaluating the quality of the graft, and minimize (or even eliminate) the possibility of criteria error in the evaluation of this important parameter. Additionally, we developed our application based on the optimization of the quality and the size of the images, in order to optimize the time required for analysis and the computational cost. Regarding the algorithms analyzed, Naïve Bayes and KNN were the best algorithms for the data set on which they were evaluated. Both displayed remarkably high levels of accuracy, sensitivity, and specificity, while also proving to be the fastest in both the training and classification steps, with minimal consumption of hardware resources.

In the future, these algorithms may be implemented in specialized and automatized image analysis applications for liver transplantation, with specific use in Sudan-stained sections.

4. Materials and Methods

4.1. Liver Samples and Histochemical Procedures

Eight micrometer-thick sections were obtained from donor liver samples ($n = 20$), which were preserved at $-80\text{ }^{\circ}\text{C}$. These sections were stained with an improved Sudan procedure, a specific histochemical fat-staining procedure routinely used in our department to evaluate fat infiltration in frozen samples. Briefly, all samples were sectioned (8

micrometers thick) and stained with a mix of Sudan III and Sudan IV dyes at 50%/50% vol. (Sigma Aldrich, Madrid, Spain) for 10 min at room temperature. The sections were Hematoxylin (Thermo, Barcelona, Spain) counterstained and finally mounted with an aqueous permanent mounting medium. Fat Sudan-positive structures were identified as intracellular orange vacuoles (Figure 5, left). The main artifacts related to the procedure that were detected were air bubbles (due to the aqueous mounting media), sectioning artifacts, unspecific hematoxylin deposits, and unstained blank spaces, mainly due to sinusoid dilation, vessels, and hepatocyte ballonization unrelated to fat deposits.

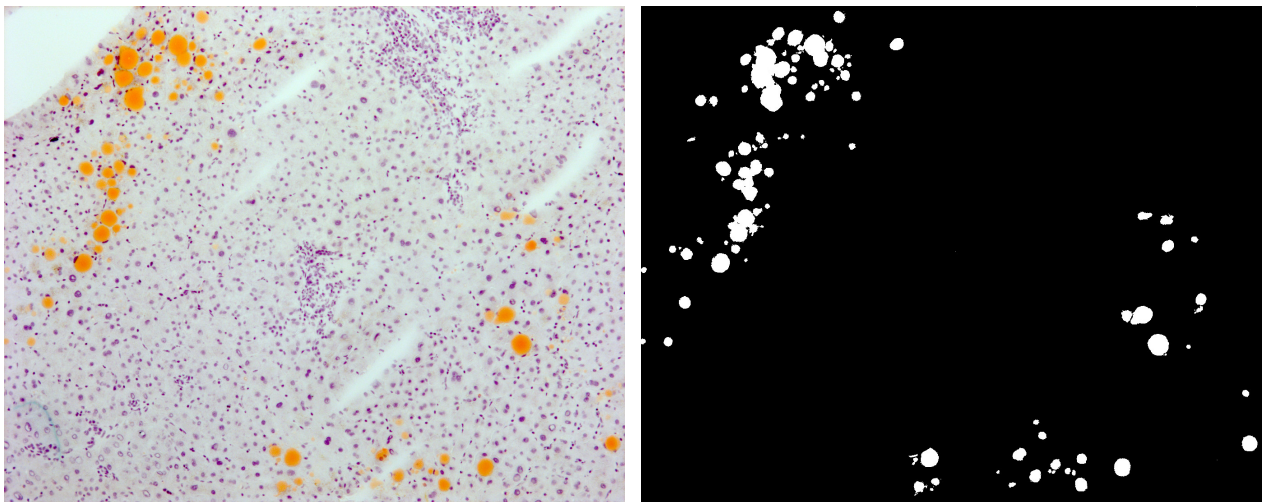


Figure 5. Original image (left) and manually classified image (right).

4.2. Imaging

A number of digital images were obtained from stained samples by using a direct-light microscope (Zeiss Axio A10, Carl Zeiss, Jena, Germany) equipped with a high-quality digital camera (Axio Cam 506, Zeiss) and specialized software (Zeiss Zen ver. 3.0). The technical specifications of the camera are detailed in Table 4 (the complete specifications can be found at <https://www.zeiss.com/microscopy/int/products/microscope-cameras.html>, accessed on 11 March 2021). Histologically, macrovesicular steatosis is defined as the presence of a large-sized intracellular fat vacuole (with a minimum of two-times bigger than normal nucleus size) which displaces the nucleus to the periphery of the cells [23,24]. To determine the standard diameter of the normal hepatocyte nucleus in frozen sections, we measured 3000 hepatocyte nuclei in frozen sections from 10 healthy livers only stained with hematoxylin, measuring the media of all sizes ($73.356 \pm 0.177 \mu\text{m}^2$) and established this value as a standard. Therefore, we considered intracellular macrovesicular fat vacuoles as those with a size $\geq 146.71 \mu\text{m}^2$.

Table 4. Zeiss Axiocam basic specifications.

Sensor Model	Sony ICX 694, EXview HAD CCD II
Sensor pixel count	6 Megapixel. 2752 (H) \times 2208 (V)
Pixel size	4.54 μm \times 4.54 μm
Exposure time range	250 μs to 60 s.
Spectral sensitivity	Aprox. 400–720 nm. RGB Bayer color filter mask

4.3. Generating Learning Models

To carry out the classification of images, several learning models were generated using different machine learning and deep learning algorithms. Specifically, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), simple Neural Network (NN), and neural network with Tensorflow and Keras [25] (using GPU) were used. Each algorithm, except for Keras, was evaluated with the default param-

eters, and the image classification process was parallelized. Regarding Keras, a densely connected first layer with 6 nodes with *sigmoid* activation and a final layer with 2 nodes and *softmax* activation were used. *Adamax* was used as the network optimizer. In addition, the computer graphics were configured to use the GPU for training and image classification with this algorithm.

To generate the models, as well as to evaluate their performance, images from optical microscopy of liver tissues at 100, 200, and 400× magnifications were used. Images were obtained with exposure, brightness, and contrast values self-adjusted by the camera software, and images with different levels of adjustment were modified manually. The color histograms of the images were adjusted to the histogram of a reference image defined as the best-fitting by a match-histogram algorithm.

Twenty images with 1920 × 1080 pixel resolution at different magnifications were used. For each image, windows of 10 × 10, 20 × 20, 50 × 50, and 100 × 100 pixels were manually extracted, depending on the enlargement level and size of the vacuoles. A total of 10 50 × 50 windows, 10 100 × 100 windows, 50 20 × 20 windows, and 50 10 × 10 windows were finally taken. The total number of pixels obtained was 200,000. For every pixel, a 6-characteristic vector, defined by RGB and CIE L*a*b color spaces, was obtained:

$$FV_n = \begin{bmatrix} R_i & G_i & B_i & L_i & a_i & b_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_n & G_n & B_n & L_n & a_n & b_n \end{bmatrix} \forall i = 1, \dots, n, \quad (1)$$

where FV_n is the set of feature vectors for all pixels (n) used to construct the models, and $R_i, G_i, B_i, L_i, a_i, b_i$ are the values of the red (R), green (G), blue (B), lightness (L^*), green to red (a^*), and blue to yellow (b^*) channels, respectively, corresponding to the i th pixel. This type of feature vector has already been successfully tested in other works related to image analysis [22]. Pixels were tagged with 1 or 0, depending on whether they belonged to a region of the image where there was a fat vacuole (1) or not (0).

Of the total number of pixels, randomized subsets of 100,000, 50,000, 10,000, 5000, and 1000 pixels were selected to train different models with different numbers of pixels. From each pixel subset, 70% were used to train the models, while the remaining 30% were used for testing. The data splitting was carried out by stratified random sampling, in order to obtain a proportionate number of pixels from each class.

Finally, the performance of each algorithm was assessed using the test data sets (30% remaining) of the corresponding training sets. The data splitting, training, and testing processes were executed 10 times with each algorithm and with every data set, in order to determine the average time each algorithm took to train and the average time to classify. Likewise, the AUC of each algorithm for every subset of data was calculated by ROC curves.

4.4. Classification Time

With the trained models at 50,000 pixels, the same image was classified at two different resolutions—2752 × 2208 and 1376 × 1104 pixels—in order to determine the time spent by the different models in classifying each image with a different number of threads.

Every performance test was carried out on a laptop with an Intel Core i7-9750H processor at 2.6 GHz, with 6 cores and 12 threads, 16 Gb RAM, and a 4 Gb NVIDIA GTX1620 graphics card.

4.5. Classification Validation

The automatically classified images were compared with the same manually classified images by two expert pathologists (Figure 5) through a simple binary image subtraction operation, hence obtaining the TP, FP, TN, and FN values, with respect to the reference image, in order to obtain the overall accuracy, sensitivity, specificity, and precision scores (Equations (2)–(5), respectively). We defined those pixels classified as fat vacuole matching the manually classified image as the TP, those classified as non-vacuole matching the

manual image as the TN and, finally, those wrongly classified in the fat vacuole or non-vacuole categories as FP and FN, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

As, in many cases, the fat vacuoles are totally united in the image, to correctly quantify the size and number of vacuoles, it is necessary to have an automatic mechanism that, as far as possible, distinguishes two or more vacuoles together as independent entities. To do this, morphological segmentation has been applied using the watershed algorithm, which has been widely used in the analysis of biomedical and biological images for cell segmentation [26–28]. In those cases of extremely high values of ME infiltration (e.g., 70–80%), this classification algorithm experienced some problems in classifying extremely large overlapped vacuoles, although the final result of ME value was not altered due to this limitation.

4.6. Web Application Development

In order to assist in the assessment of ME degree, a web application was developed using the *Shiny* [29] framework provided by *R* [30], which enables the rapid development of web applications and simplifies the integration of additional programming languages. The web application was developed in such a way that the user may sequentially follow the steps that lead them from image uploading to the degree of ME quantification.

The user sets the number of microscope magnifications when capturing the image (Figure 6(1)). This determines the size of each pixel (in microns). Then, after loading the image (Figure 6(2)), the application allows the user to select a pre-trained model from several algorithms, or to manually train the model by marking the points of interest on the image (Figure 6(3),(4)). Afterwards, the application classifies the image and returns the number and extension of fat vacuoles over the total image, as well as the macrovesicle percentage (Figure 7).

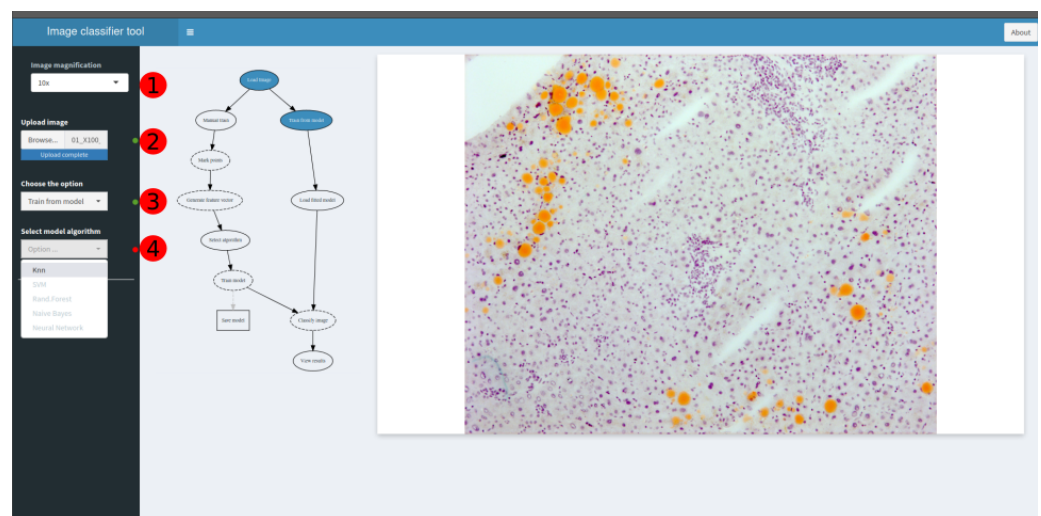


Figure 6. Web application interface image classification steps: (1) Objective magnification selector; (2) Image uploader; (3) Manual or pre-trained model selector; and (4) (if pre-trained) algorithm selector.

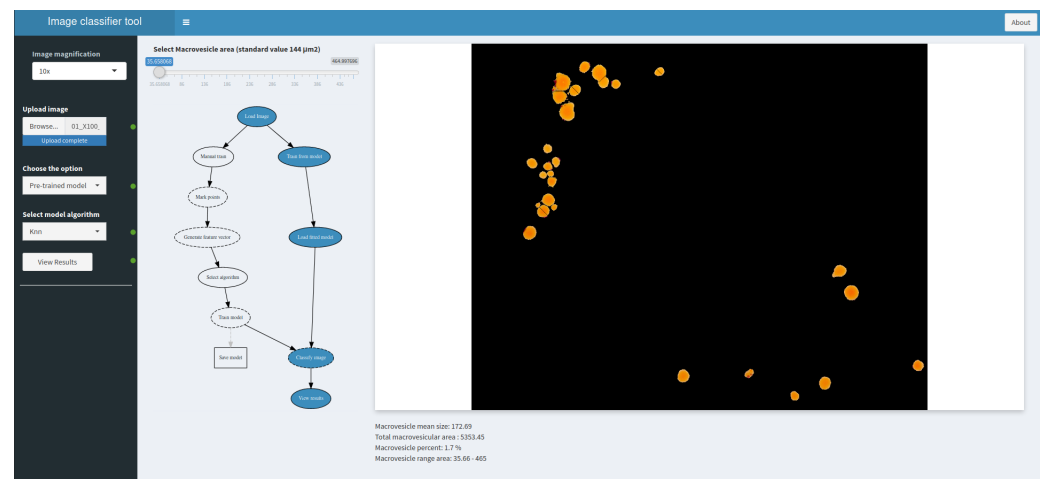


Figure 7. Result of image classification and quantification of fatty vacuoles.

The whole image classification process was conducted in python, mainly using the *scikit-image* [31], *scikit-learn* [32], and *Keras* libraries. Its integration with R was carried out using the R *Reticulate* library [33], which allows for the execution of python code inside R applications.

5. Conclusions

Sudan staining is a suitable stain procedure, which can be used in the evaluation of macrovesicular steatosis of the graft in pre-liver transplantation histopathological evaluation. It is a quick and easy technique that, unlike the hematoxylin and eosin stain, is specific to fat identification. Due to its specificity, this stain is suitable for automatic and quantitative evaluation by the use of machine learning algorithms.

The machine learning algorithms Naïve Bayes and KNN showed the best results, in terms of speed and accuracy, in all tests performed for the automatic identification of macrovesicular steatosis in Sudan pre-transplant liver stained sections.

Therefore, the automatic evaluation of macrovesicular steatosis may be performed during the histopathologic evaluation of the quality of the liver graft in the pre-transplant evaluation by using Sudan-stained sections, while other parameters can be established by direct examination of hematoxylin and eosin stained sections by an expert pathologist.

Author Contributions: Conceptualization, F.P.-S., C.M.C., P.R., and E.M.-B.; Investigation, F.P.-S., M.R.-P., A.S.N., and M.C.-R.; Formal analysis, F.P.S.; Software F.P.-S., A.E.-G., M.A.P.-G., and M.D.C.L.-G.; Validation J.d.I.P.-M., A.S.N., E.M.-B., and C.M.M.; Writing—original draft C.M.M., F.P.-S., M.R.-P., and P.R.; Writing—review and editing A.E.-G., M.D.C.L.-G., M.A.P.-G., C.M.M., F.P.-S., M.R.-P., and M.C.-R.; Funding acquisition, A.E.-G. and C.M.M.; Methodology C.M.M. and F.P.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Seneca foundation (Agencia de Ciencia y Tecnología de la Región de Murcia); Grant number: 21104/PDC/19.

Institutional Review Board Statement: The Ethics Advisory Board (EAB) did not consider appropriate to ethically this technological innovation research project due to the fact that it does not use biological samples from specific patients. Additionally, this is an innovative research project which is developed prior to its clinical application. However, it has been positively evaluated by the Internal Scientific Committee of our institution (IMIB).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data example is available at <https://github.com/MiriamRiquelmeP/Rython.git>, accessed on 11 March 2021.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Sample Availability: Software is available at <https://github.com/MiriamRiquelmeP/Rython.git>, accessed on 11 March 2021.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area Under the Curve
H&E	Hematoxylin and Eosin
KNN	K-nearest neighbors
ME	Macrovesicular steatosis
NB	Naïve Bayes
NN	Neural Network
RF	Random Forest
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine

References

- Gurakar, A.; Tasdogan, B.E.; Akosman, S.; Gurakar, M.; Simsek, C. Update on Liver Transplantation: What is New Recently? *Euroasian J. Hepato-Gastroenterol.* **2019**, *9*, 34–39. [[CrossRef](#)] [[PubMed](#)]
- Pan, E.T.; Yoeli, D.; Galvan, N.T.N.; Kueht, M.L.; Cotton, R.T.; O'Mahony, C.A.; Goss, J.A.; Rana, A. Cold ischemia time is an important risk factor for post-liver transplant prolonged length of stay. *Liver Transplant.* **2018**, *24*, 762–768. [[CrossRef](#)] [[PubMed](#)]
- Fiorentino, M.; Vasuri, F.; Ravaioli, M.; Ridolfi, L.; Grigioni, W.F.; Pinna, A.D.; D'Errico-Grigioni, A. Predictive value of frozen-section analysis in the histological assessment of steatosis before liver transplantation. *Liver Transplant.* **2009**, *15*, 1821–1825. [[CrossRef](#)] [[PubMed](#)]
- Chu, M.J.; Dare, A.J.; Phillips, A.R.; Bartlett, A.S. Donor Hepatic Steatosis and Outcome After Liver Transplantation: A Systematic Review. *J. Gastrointest. Surg.* **2015**, *19*, 1713–1724. [[CrossRef](#)] [[PubMed](#)]
- Chavin, K.D.; Taber, D.J.; Norcross, M.; Pilch, N.A.; Crego, H.; Mcgillicuddy, J.W.; Bratton, C.F.; Lin, A.; Baliga, P.K. Safe use of highly steatotic livers by utilizing a donor/recipient clinical algorithm. *Clin. Transplant.* **2013**, *27*, 732–741. [[CrossRef](#)] [[PubMed](#)]
- Choi, W.T.; Jen, K.Y.; Wang, D.; Tavakol, M.; Roberts, J.P.; Gill, R.M. Donor Liver Small Droplet Macrovesicular Steatosis is Associated with Increased Risk for Recipient Allograft Rejection. *Am. J. Surg. Pathol.* **2017**, *41*, 365–373. [[CrossRef](#)] [[PubMed](#)]
- McCormack, L.; Dutkowski, P.; El-Badry, A.M.; Clavien, P.A. Liver transplantation using fatty livers: Always feasible? *J. Hepatol.* **2011**, *54*, 1055–1062. [[CrossRef](#)] [[PubMed](#)]
- El-Badry, A.M.; Breitenstein, S.; Jochum, W.; Washington, K.; Paradis, V.; Rubbia-Brandt, L.; Puhan, M.A.; Slankamenac, K.; Graf, R.; Clavien, P.A. Assessment of hepatic steatosis by expert pathologists: The end of a gold standard. *Ann. Surg.* **2009**, *250*, 691–696. [[CrossRef](#)]
- Yersiz, H.; Lee, C.; Kaldas, F.M.; Hong, J.C.; Rana, A.; Schnickel, G.T.; Wertheim, J.A.; Zarrinpar, A.; Agopian, V.G.; Gornbein, J.; et al. Assessment of hepatic steatosis by transplant surgeon and expert pathologist: A prospective, double-blind evaluation of 201 donor livers. *Liver Transplant.* **2013**, *19*, 437–449. [[CrossRef](#)]
- Sun, L.; Marsh, J.N.; Matlock, M.K.; Chen, L.; Gaut, J.P.; Brunt, E.M.; Swamidass, S.J.; Liu, T.C. Deep learning quantification of percent steatosis in donor liver biopsy frozen sections. *EBioMedicine* **2020**, *60*, 103029. [[CrossRef](#)]
- Li, M.; Song, J.; Mirkov, S.; Xiao, S.Y.; Hart, J.; Liu, W. Comparing morphometric, biochemical, and visual measurements of macrovesicular steatosis of liver. *Hum. Pathol.* **2011**, *42*, 356–360. [[CrossRef](#)]
- Marsman, H.; Matsushita, T.; Dierkhising, R.; Kremers, W.; Rosen, C.; Burgart, L.; Nyberg, S.L. Assessment of Donor Liver Steatosis: Pathologist or Automated Software? *Hum. Pathol.* **2004**, *35*, 430–435. [[CrossRef](#)]
- Boyles, T.H.; Johnson, S.; Garrahan, N.; Freedman, A.R.; Williams, G.T. A validated method for quantifying macrovesicular hepatic steatosis in chronic hepatitis C. *Anal. Quant. Cytol. Histol.* **2007**, *29*, 244–250.
- Forlano, R.; Mullish, B.H.; Giannakeas, N.; Maurice, J.B.; Angkathunyakul, N.; Lloyd, J.; Tzallas, A.T.; Tsiouras, M.; Yee, M.; Thursz, M.R.; et al. High-Throughput, Machine Learning-Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease. *Clin. Gastroenterol. Hepatol.* **2020**, *18*, 2081–2090. [[CrossRef](#)] [[PubMed](#)]
- García Ureña, M.A.; Ruiz-Delgado, F.C.; Moreno González, E.; Jiménez Romero, C.; García García, I.; Loinzaz Seguro, C.; González-Pinto, I.; Gómez Sanz, R. Hepatic steatosis in liver transplant donors: Common feature of donor population? *World J. Surg.* **1998**, *22*, 837–844. [[CrossRef](#)]

16. Guo, X.; Wang, F.; Teodoro, G.; Farris, A.B.; Kong, J. Liver steatosis segmentation with deep learning methods. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 24–27. [[CrossRef](#)]
17. Roy, M.; Wang, F.; Vo, H.; Teng, D.; Teodoro, G.; Farris, A.B.; Castillo-Leon, E.; Vos, M.B.; Kong, J. Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Lab. Investig.* **2020**, *100*, 1367–1383. [[CrossRef](#)] [[PubMed](#)]
18. Vanderbeck, S.; Bockhorst, J.; Komorowski, R.; Kleiner, D.E.; Gawrieh, S. Automatic classification of white regions in liver biopsies by supervised machine learning. *Hum. Pathol.* **2014**, *45*, 785–792. [[CrossRef](#)] [[PubMed](#)]
19. Moccia, S.; Mattos, L.S.; Patrini, I.; Ruperti, M.; Poté, N.; Dondero, F.; Cauchy, F.; Sepulveda, A.; Soubrane, O.; De Momi, E.; et al. Computer-assisted liver graft steatosis assessment via learning-based texture analysis. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 1357–1367. [[CrossRef](#)] [[PubMed](#)]
20. Moosburner, S.; Gassner, J.M.; Nösser, M.; Pohl, J.; Wyrwal, D.; Claussen, F.; Ritschl, P.V.; Dragun, D.; Pratschke, J.; Sauer, I.M.; et al. Prevalence of steatosis hepatitis in the eurotransplant region: Impact on graft acceptance rates. *HPB Surg.* **2018**, *2018*. [[CrossRef](#)]
21. Panqueva, R.d.P.L. Histopathological Evaluation of Liver Donors: An Approach to Intraoperative Consultation during Liver Transplantation. *Rev. Colomb. Gastroenterol.* **2015**, *30*, 485–495.
22. Navarro, P.J.; Pérez, F.; Weiss, J.; Egea-Cortines, M. Machine learning and computer vision system for phenotype data acquisition and analysis in plants. *Sensors* **2016**, *16*, 641. [[CrossRef](#)]
23. Haas, M. Histopathology of liver transplantation. In *Transplantation Pathology*, 2th ed.; Cambridge Medicine Press: Cambridge, UK, 2018; Volume 18, p. 67. [[CrossRef](#)]
24. Lefkowitz, J.H. (Ed.) Steatosis, steatohepatitis and related conditions. In *Scheuer's Liver Biopsy Interpretation*, 9th ed.; Elsevier: Amsterdam, The Netherland, 2015; p. 440.
25. Chollet, F. Keras. 2015. Available online: <https://keras.io> (accessed on 11 March 2021).
26. Maity, M.; Jaiswal, A.; Gantait, K.; Chatterjee, J.; Mukherjee, A. Quantification of malaria parasitaemia using trainable semantic segmentation and capsnet. *Pattern Recognit. Lett.* **2020**, *138*, 88–94. [[CrossRef](#)]
27. Koyuncu, C.F.; Arslan, S.; Durmaz, I.; Cetin-Atalay, R.; Gunduz-Demir, C. Smart Markers for Watershed-Based Cell Segmentation. *PLoS ONE* **2012**, *7*, e48664. [[CrossRef](#)] [[PubMed](#)]
28. Tek, F.B.; Dempster, A.G.; Kale, I. Blood Cell Segmentation Using Minimum Area Watershed and Circle Radon Transformations. In *Mathematical Morphology: 40 Years on*; Springer: Dordrecht, The Netherlands, 2005; pp. 441–454. [40](#). [[CrossRef](#)]
29. Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. Package 'Shiny': Web Application Framework for R. 2021. Available online: <https://cran.r-project.org/web/packages/shiny> (accessed on 11 March 2021).
30. RDC Team. A Language and Environment for Statistical Computing. 2018. Volume 2. Available online: <https://www.R-project.org> (accessed on 11 March 2021).
31. Van Der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-image: Image processing in python. *PeerJ* **2014**, *2014*, e453. [[CrossRef](#)] [[PubMed](#)]
32. Pedregosa, F.; Grisel, O.; Weiss, R.; Passos, A.; Brucher, M.; Varoquax, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. Ushey, K.; Allaire, J.J.; Tang, Y. Reticulate: Interface to 'Python'. 2020. Available online: <https://cran.r-project.org/web/packages/reticulate> (accessed on 11 March 2021).