

Journal Pre-proof

Automated image analysis method to detect and quantify fat cell infiltration in hematoxylin and eosin stained human pancreas histology images

Roshan Ratnakar Naik , Annie Rajan , Nehal Kalita

PII: S2667-1603(23)00013-3
DOI: <https://doi.org/10.1016/j.bbadv.2023.100084>
Reference: BBADVA 100084



To appear in: *BBA Advances*

Received date: 6 November 2022
Revised date: 14 February 2023
Accepted date: 15 February 2023

Please cite this article as: Roshan Ratnakar Naik , Annie Rajan , Nehal Kalita , Automated image analysis method to detect and quantify fat cell infiltration in hematoxylin and eosin stained human pancreas histology images, *BBA Advances* (2023), doi: <https://doi.org/10.1016/j.bbadv.2023.100084>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Highlights

- We have developed an automated tool, Fatquant, for identification of fat cells based on its diameter in hematoxylin and eosin tissue sections such as pancreas, liver which can aid the pathologist for diagnosis of fatty pancreas and related metabolic conditions. Fatquant is unique as current fat automated tools (adiposoft, adipocount) works well for homogeneous white fat tissue but not for other tissue samples. The currently available liver, pancreas analysis tool are either deep learning based, require training datasets unlike Fatquant or are not open source software.

Title

Automated image analysis method to detect and quantify fat cell infiltration in hematoxylin and eosin stained human pancreas histology images

Authors

Roshan Ratnakar Naik¹, Annie Rajan², Nehal Kalita³

Affiliations

¹ Dr. Roshan Ratnakar Naik, Assistant Professor, Department of Biotechnology, Parvatibai Chowgule College of Arts & Science, Margao-Goa, 403601. ORCID ID: 0000-0003-4515-871X.
Email address: ros007swa@gmail.com

² Mrs. Annie Rajan, Associate Professor, Department of Computer Science and Technology,

Dhempe College of Arts and Science, Miramar, Panaji-Goa, 403 001. ORCID ID: 0000-0001-6201-2654. Email address: ann_raj_2000@yahoo.com

³ Nehal Kalita, M.Sc. Computer Science. ORCID ID: 0000-0001-6997-5621. Email address: nehalkalita94@gmail.com

Corresponding Author: Dr. Roshan R. Naik

Author contributions

R.R.N. conceptualized the idea and wrote the manuscript.

A.R. conceptualized the idea.

N.K. formulated the mathematical approach and wrote the Python programs.

Abstract

Fatty infiltration in pancreas leading to steatosis is a major risk factor in pancreas transplantation. Hematoxylin and eosin (H and E) is one of the common histological staining techniques that provides information on the tissue cytoarchitecture. Adipose (fat) cells accumulation in pancreas has been shown to impact beta cell survival, its endocrine function and pancreatic steatosis and can cause non-alcoholic fatty pancreas disease (NAFPD). The current automated tools (E.g. Adiposoft) available for fat analysis are suited for white fat tissue which is homogeneous and easier to segment unlike heterogeneous tissues such as pancreas where fat cells continue to play critical physiopathological functions. The currently, available pancreas segmentation tool focuses on endocrine islet segmentation based on cell nuclei detection for diagnosis of pancreatic cancer. In the current study, we present a fat quantifying tool, Fatquant, which identifies fat cells in heterogeneous H and E tissue sections with reference to diameter of fat cell. Using histological images from a public database, we observed an intersection over union of 0.797 to 0.962 and 0.675 to 0.937 for manual versus Fatquant analysis of pancreas and liver, respectively.

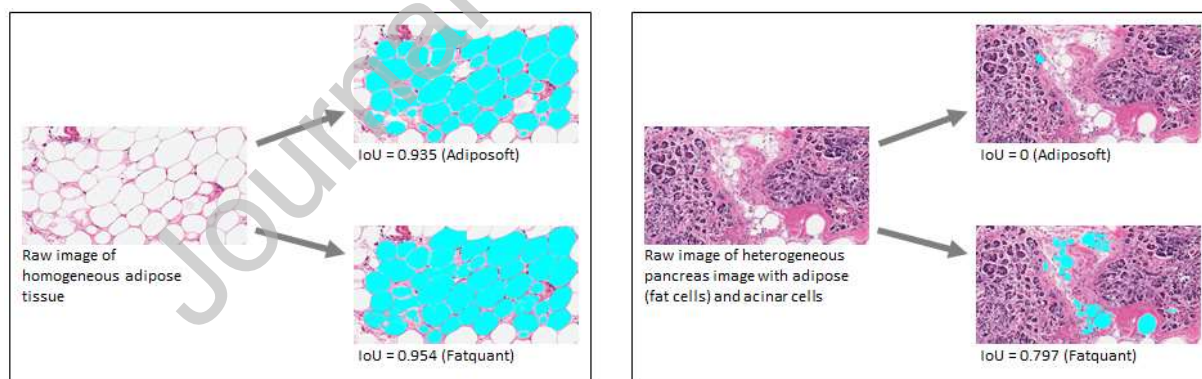
Keywords

Adipose or fat cells, Hematoxylin and Eosin, image processing, tissue segmentation, pancreas, liver

Highlights

We have developed an automated open source software Fatquant, for identification of fat cells based on its diameter in hematoxylin and eosin tissue sections such as pancreas, liver which can aid the pathologist for diagnosis of fatty pancreas and related metabolic conditions. Fatquant is unique as current fat automated tools (adiposoft, adipocount) works well for homogeneous white fat tissue but not for other tissue samples. The currently available liver, pancreas analysis tool is either deep learning based, require training datasets unlike Fatquant or are not open source software.

Graphical Abstract



Currently available fat quantification tools like Adiposoft can analyze homogenous fat tissue (left) with intersection over union (IoU) of 0.935 and 0.954 with Adiposoft and Fatquant, respectively. While in heterogenous tissue (e.g. pancreas on right) which contains fat cells, acinar cells, Adiposoft fails to detect fat cells with $\text{IoU}=0$ while Fatquant had $\text{IoU}=0.797$.

1. Introduction

The accumulation of fats especially in the abdominal area causes insulin resistance, deposition of fats (steatosis), inflammation, fibrosis in the pancreas leading to non-alcoholic fatty pancreatic disease (NAFPD). The consequence of pancreatic fat infiltration might provoke a decrease in endocrine (β -cell) number, function leading to more rapid progression to diabetes with NAFPD suggested as an early marker of glucometabolic disturbance [1]. Whole or islet pancreas transplantation simultaneously with kidney or after kidney transplant in comorbidity are some of the known methods to treat adult type 1 diabetes (T1D) [2], however, surgical, immunosuppression [2] and fat infiltration in pancreas are some of the risk factors that can affect the clinical outcome [3,4]. Fatty pancreas has a prevalence of 35% and may lead to pancreatitis, diabetes mellitus or pancreatic cancer [5,6].

Pancreatic steatosis can be diagnosed on ultrasound, computed tomography (CT) scan or magnetic resonance imaging (MRI) but pancreatic biopsy remains best method to detect pancreatic fat concentration [5,7]. The histological and MRI tools exhibit good agreement in detecting fat in pancreas [8,9] but the latter cannot show fat accumulation at the cellular level and is expensive. Hematoxylin and eosin (H & E) is a widely used histological tissue staining technique for medical diagnosis and scientific research. Hematoxylin stains cell nuclei blue while eosin stains the cytoplasm and connective tissue pink thus allowing microscopic differentiation of tissue cytoarchitecture in sections. The analysis involves manual examination by pathologist to ascertain presence/absence of disease markers and or grading of disease progression which is semi-quantitative and subjective in nature [10].

To complement the current manual assessment, several digital tools have been developed such as Adiposoft [11], and AdipoCount [12] which works well for homogeneous white fat tissue (Figure 1). However, they cannot efficiently identify fat cells in H and E tissues such as pancreas or liver due to presence of a mixture of various cells and structures complicating the fat boundary detection which is one the earliest steps in cell segmentation [13]. Such tool developed based on

nuclear displacement and lipid droplet size analysis have been reported for automated analysis of fat cell infiltration (steatosis) in H and E liver images [14]. Unlike liver tissue, where fats accumulate in hepatocytes, in pancreas fat infiltration and deposition occurs both in acinar and islet cells [15].

Usually, these algorithms involve splitting the image into various color channels with the red channel binarized using automatic thresholding method to separate the bright pink fat areas from dark purple-bluish cell nuclei. Subsequently, a watershed algorithm is applied to fillup missing fat cell membrane and improves cell count [11,12]. The output of these processes includes the labels and statistical analysis of individual cells. These tools have been successfully applied to white fat tissue, however, other organs like liver, pancreas, lungs have been challenging due to heterogeneous cell types (Figure 2).



Figure 1-Homogenous fat tissue

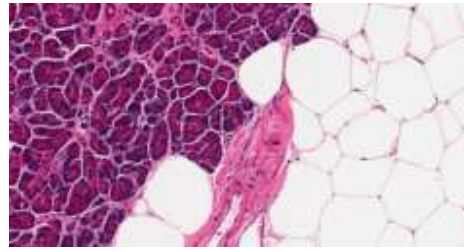


Figure 2-Heterogenous pancreatic tissue with fat and acinar cells

Pancreas is a heterogenous tissue and manual analysis of regions is a tedious process that lacks reproducibility [16]. The existing methods of pancreatic islet segmentation depends on cell nuclei detection, then a classifier is applied to recognize different cell types but is based on assumption that islets have high density cells [17,18]. Recent studies have utilized a supervised

learning framework for islet segmentation in H & E stained pancreatic images to partition images into superpixels and extract color-texture features, process them, and finally a linear support vector machine is trained and applied to segment testing images [19]. Moreover, QuPath software allows identification of endocrine islet cells in immunofluorescent images of pancreas tissue [16]. In the current study, we have developed an automated tool, Fatquant, in which the fat cells were identified in processed images by calculating the diagonal of a square circumscribed by circle.

2. Materials and Methods

The H and E images for the analysis were downloaded from the Genotype-Tissue Expression (GTEx) public portal (Broad Institute, Cambridge, MA, USA) using the Histology Viewer tool [20]. The GTEx tissue image library contains high-resolution histology images for various tissue types from several postmortem donors. Spherical or oval white spaces were categorized as fat cells while large and irregular white spaces were grouped as artifacts. A sample size of ten each pancreas histology images with subject IDs GTEx-11DXZ-0826, GTEx-1122O-0726, GTEx-1117F-1726, GTEx-117YW-0926, GTEx-13PVQ-2026, GTEx-13FHP-1926 and GTEx-11WQC-0926 and liver histology images with subject ID 11EMC-0326, 11EQ9-0526, 11GS4-0926, 11GSO-0826, 11NV4-1326, 11O72-0926 were analyzed. Each subject ID denotes unique individuals (7 for pancreas, 6 for liver), while remaining were random images with varying fat percentage and distinctive neighboring cell. The images have magnification of 20x (0.4942 mpp—microns per pixel) [21]. The sample images were taken using the Snapshot tool of Aperio ImageScope software. The source code was written in Python version 3.8.0 (Python Software Foundation, Beaverton, Oregon, USA) with Core i7 3rd Generation CPU, 12 GB RAM and Intel HD graphics 4000 GPU.

2.1. Data and code availability

All the images, annotations along with relevant data code can be found at the following GitHub repository: <https://github.com/anniedhempe/Fatquant>. The procedure to run this tool is mentioned in the Readme file. In addition, we have provided a GUI version of Fatquant tool along with tutorial at the repository: <https://github.com/nehalkalita/Fatquant-GUI>

2.2. Image processing

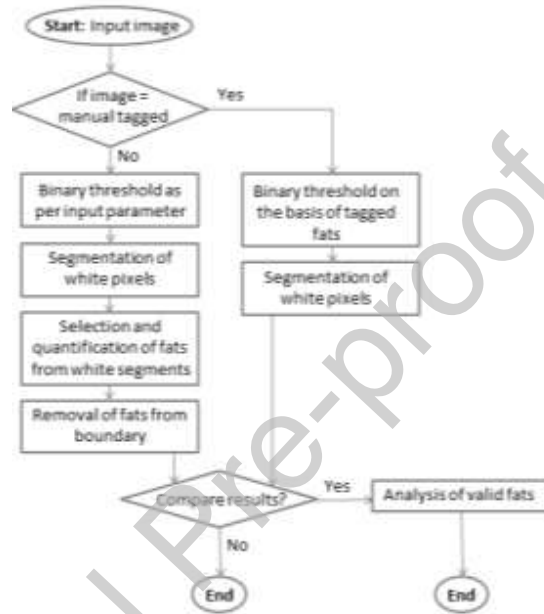


Figure 3: Flowchart depicting the image processing.

The procedure for image processing is briefly demonstrated in the flow chart shown in Figure 3.

The steps used are elaborated below.

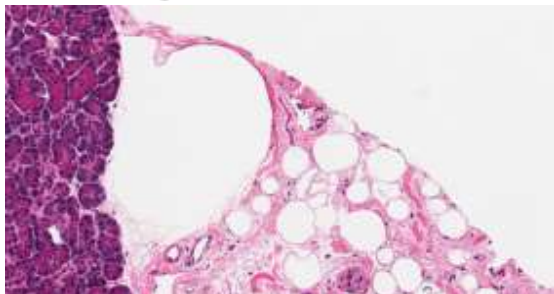


Figure 4

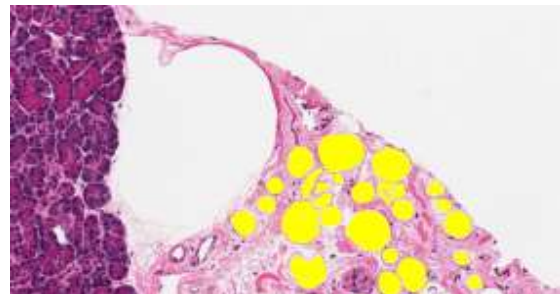


Figure 5

Figure 4 is a sample raw image of dimension 1716 x 905 pixels from subject ID GTEX-11DXZ-

0826. Analysis on this image is referred while explaining the image processing steps. Figure 5 is an altered form of the image in Figure 4 where the valid fat cells are manually tagged with yellow color (RGB: 255, 255, 0).

2.2.1. Binary thresholding of input image

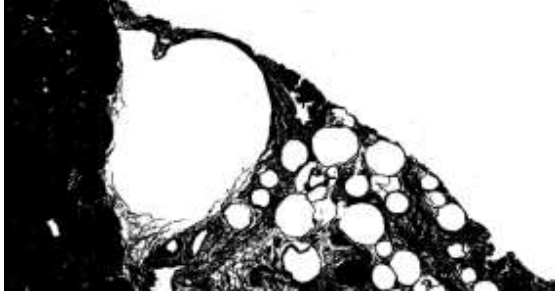


Figure 6

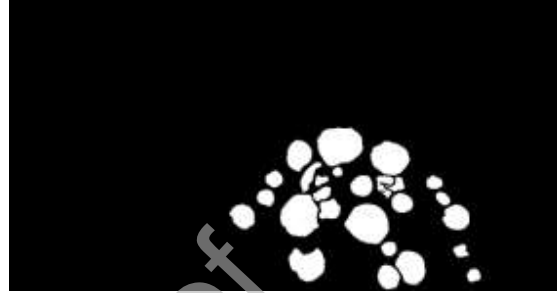


Figure 7

The color of fat cells in images from GTEX portal ranges approximately between 225 to 255 grayscale values. There can also be other parts of pancreas which has the same range of color. But applying binary threshold on an image can help in getting rid of many unwanted parts. The pixels of an image whose color values are at least equal to the input parameter value of threshold (e.g. 227) are taken into consideration for further processing and are assigned a new grayscale value 255. The other pixels are assigned value 0. Figure 6 is a thresholded image of Figure 4 with parameter value 230. Figure 7 is also a thresholded image but of Figure 5 where pixels representing the tagged fat cells are assigned grayscale value 255 and the rest is assigned value 0.

2.2.2. Segmentation of white pixels from thresholded image

White pixels are initially segmented by combining tile rendering with scanline rendering and then identifying possible merge of segments in a tile with their immediate neighbors. Tile rendering has been implemented in this system as it helps in reducing time complexity for segments covering large area.

Processing time for segmentation was tested with four sizes of square tiles, which were of length 35 pixels (processing time: 24.23 seconds), 50 pixels (processing time: 15.32 seconds), 70 pixels

(processing time: 12.88 seconds) and 100 pixels (processing time: 15.37 seconds). In this experiment, tile size of 70 pixels was used for analyzing all the images since it took the least amount of time.

After an input image is divided into multiple tiles, scanline rendering is performed within each tile to identify white pixels and form possible segments with their neighbors on left or top. In this experiment, left neighbors are given first preference. Once scanline rendering is performed till the last row of a tile then the identified segments are merged on the basis of their vertical neighbors. Segments which have only diagonal neighbors with another segment and not vertical, are not merged.

Figure 8 is a diagrammatic representation of segmentation performed on a square tile of length 15 pixels. Figure 8 (a) represents a thresholded image where segmentation is to be performed. Figure 8 (b) shows six segments created while iterating once till the last row. Figure 8 (c) shows total segments getting reduced to four due to merging.

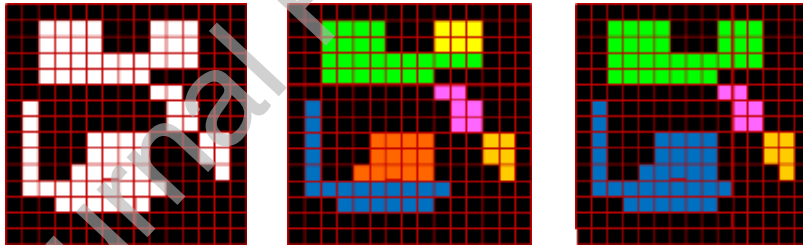


Figure 8 (a)

Figure 8 (b)

Figure 8 (c)

Figure 9 shows segments within tiles and the magnified part is one tile. Figure 10 shows fully merged segments of the thresholded image and varying colors (non-black) in it represent different segments.



Figure 9



Figure 10

2.2.3. Selection and quantification of fats from white segments

The system uses a square matrix of pixels to select probable valid segments. The matrix as per its dimension collects serially arranged group of neighbouring pixels across X, Y coordinates. Using this matrix, the system scans through a dimension of pixels for identifying fats. This dimension is the smallest dimension that encompasses all the segments. The matrix's location for scanning is only updated by one pixel (horizontally or vertically) in each iteration. So, determining the smallest dimension for traversal reduces time complexity. The dimension is determined by identifying positions of first white pixels in thresholded image from every direction i.e. top, bottom, left and right. These four positions denote maximum coverage of segments in each direction. Figure 11 is a diagrammatic representation of the method discussed where the red colored rectangle represents smallest dimension.

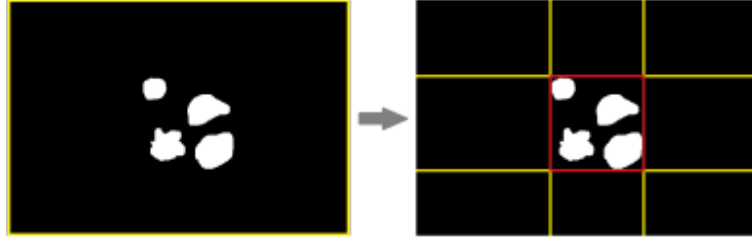


Figure 11

The square matrix is supposed to only collect pixels within the boundaries of valid segments (i.e. white pixels). The side length of that square is determined by input fat diameter values (minimum or maximum) as these diameters equates to diagonal of that square. Side length of a square, can be calculated as:

$$s = d \text{ divided by square root of } 2.$$

, where s = square side length; d = diameter. $s = \frac{d}{\sqrt{2}}$. This square can be assumed as the largest square that can get inscribed in a circle of given diameter. A set representing an elliptical shape which is equivalent of a matrix representing a square shape may be a more precise choice for identifying fats but since matrix is easier to handle so it has been chosen.

The system refers to a minimum diameter value to select segments where a square matrix having side length as per this diameter can fit somewhere in their regions. Then the system refers to a maximum diameter value to discard segments from the selected list where a matrix having side length as per this diameter can fit somewhere in their regions. This means a segment which has narrow areas in many of its portions, but has very wide area in one of its portion can also get discarded if a square matrix as per the maximum diameter can fit in that portion. Figure 12 is a diagrammatic representation of this process performed on an image of dimension 15 x 15 pixels. Figure 12 (a) has seven segments with White colored pixels out of which valid segments are to be selected. Figure 12 (b) has three segments marked with Cyan color which denotes segments

getting selected as per minimum diameter. The matrix size is of length 3 pixels. So, four segments do not get selected as the square matrix fails to fit inside the boundary of any of these segments. Figure 12 (c) has only two segments marked with Cyan color which denotes selected segment getting discarded as per maximum diameter. Here the matrix size is of length 4 pixels. So, a segment which can fit a matrix of length greater than 4 pixels is to be discarded. The previously selected segment which gets discarded could fit a matrix of length 5 pixels. Figures 13 (a) and (b) are outputs generated from the sample image with minimum (27 pixels) and maximum (130 pixels) diameter respectively. The identified fats are marked with Cyan color (RGB: 0, 255, 255).

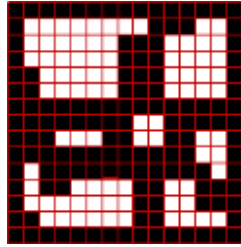


Figure 12 (a)

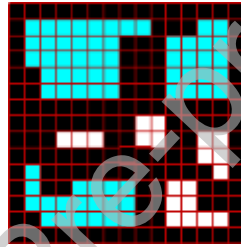


Figure 12 (b)

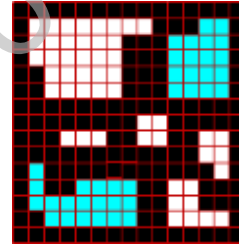


Figure 12 (c)

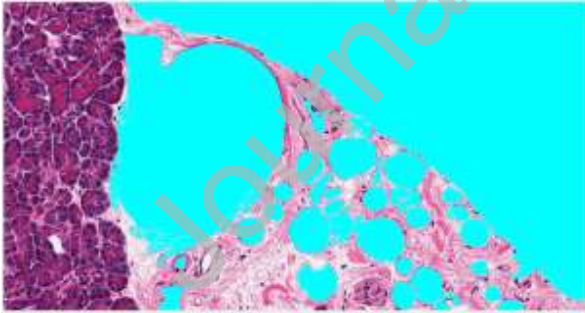


Figure 13 (a)

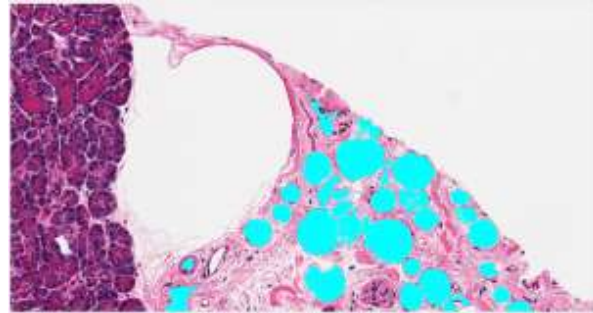


Figure 13 (b)

2.2.4. Removal of fats from boundary

Segments which are identified as fats but also contain pixels from boundary are discarded because their entire size is not known within the dimension of input image (Figure 14). While comparing Figure 14 with Figure 13 (b), it can be seen that some segments which contain pixels from bottom boundary gets discarded.

Some of these segments may even get discarded while selecting segments as per diameter. E.g. in Figure 13 (b), one big segment which has pixels in boundary gets discarded after considering maximum diameter area, whereas the segment is present in Figure 13 (a).

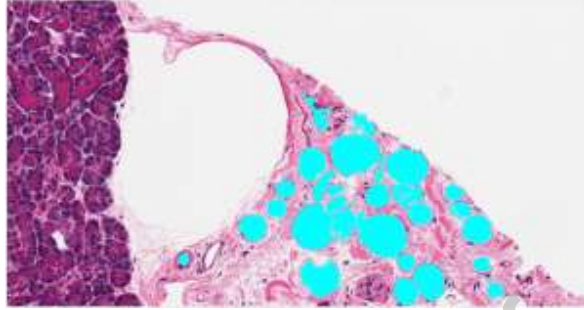


Figure 14: Fat cell removed from peripheral boundary.

Removing fats from boundaries is the final step for tagging fats using fatquant. If users do not have manual tagged data of fat cells then they can conclude the experiment after this step.

2.2.5. Analysis of valid fats

The fatquant tagged fat segments are compared with manually tagged fat segments to check the validity of our image analysis algorithm. If fatquant and manually tagged segments have pixels in common then, those pixels are considered as valid. The accuracy of the output is calculated in terms of Intersection over Union (IoU) where ‘Intersection’ means area of overlap of ground truth data with machine detected data & ‘Union’ means area of union of ground truth data and machine detected data. Its value ranges from 0 to 1, where value ‘0’ indicates no overlap and value ‘1’ indicates complete overlap. The formula used in our algorithm is:

$$IoU = \frac{TP}{FP + TP + FN}$$

, where TP = True Positive; FP = False Positive; FN = False Negative. TP is the intersection of fatquant and manually tagged pixels, FP are the pixels which are tagged by fatquant but do not become part of the intersection (i.e. Fatquant Tagged Area - TP) and FN are the pixels which are manually tagged but do not become part of the intersection (i.e. Manual Tagged Area - TP).

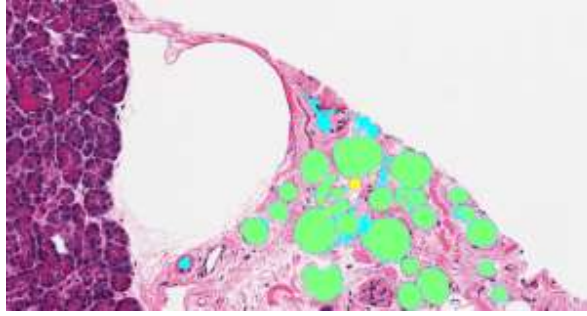


Figure 15

Figure 15 is the output generated after comparing Fatquant and manual tagged fats. The Light Green colored pixels (RGB: 127, 255, 127) represents TP, Cyan color represents FN and Yellow color represents FP. Fatquant tagged, manual tagged and TP areas are 130,816, 111,300 and 110,429 pixels respectively. Hence as per the parameters used while demonstrating the steps, the IoU value is 0.838 but it can increase if parameters are changed or a better manual tagged image is referred.

3. Results

Fat cell identification on ten sample images each of pancreas and liver dimension 1716 x 905 pixels was performed using Adiposoft and Fatquant tools. When analyzing through Adiposoft, microns per pixel value was set to 0.4942 based on their known 20x magnification [21]. The outputs are shown below.



Figure 16 (a)-Raw image

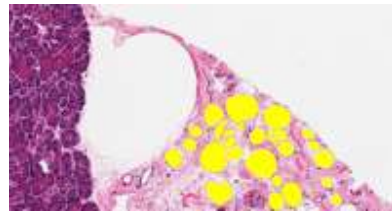


Figure 16 (b)-Tagged image



Figure 16 (c)-Adiposoft

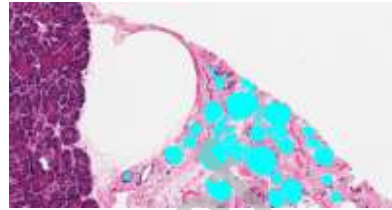


Figure 16 (d)-Fatquant

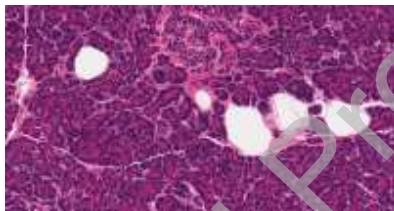


Figure 17 (a)- Raw image

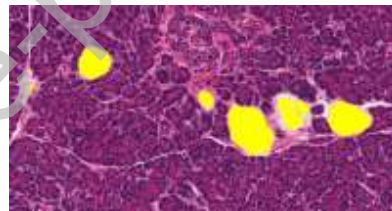


Figure 17 (b)- Tagged image



Figure 17 (c)- Adiposoft

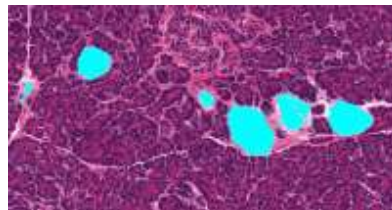


Figure 17 (d)-Fatquant

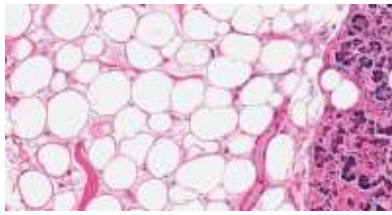


Figure 18 (a)- Raw image

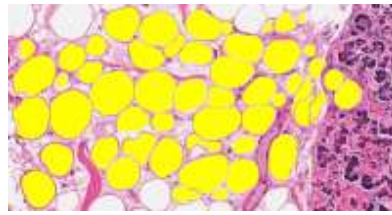


Figure 18 (b)- Tagged image

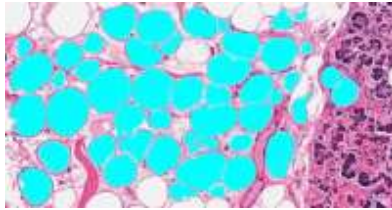


Figure 18 (c)- Adiposoft

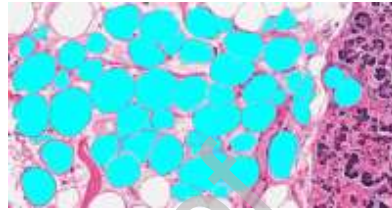


Figure 18 (d)-Fatquant



Figure 19 (a)- Raw image

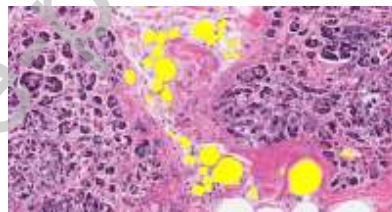


Figure 19 (b)- Tagged image



Figure 19 (c)- Adiposoft

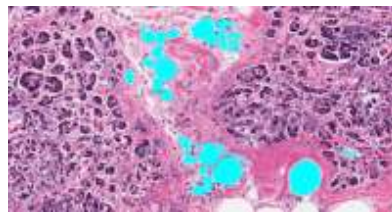


Figure 19 (d)-Fatquant



Figure 20 (a)- Raw image



Figure 20 (b)- Tagged image



Figure 20 (c)- Adiposoft



Figure 20 (d)- Fatquant

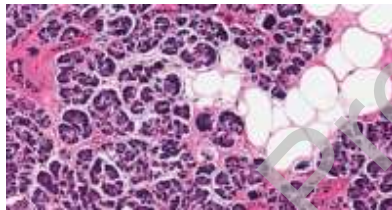


Figure 21 (a)- Raw image

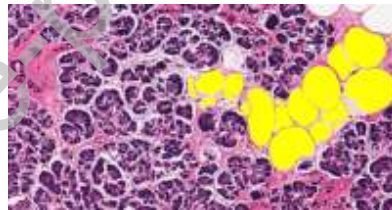


Figure 21 (b)- Tagged image

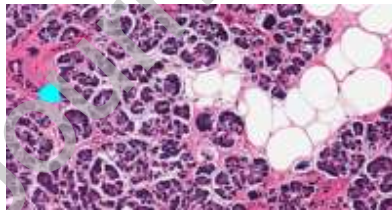


Figure 21 (c)- Adiposoft

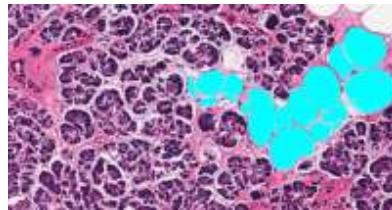


Figure 21 (d)- Fatquant

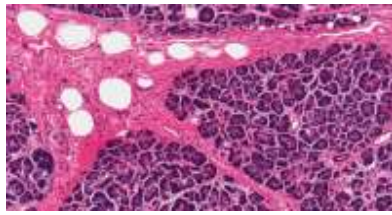


Figure 22 (a)- Raw image

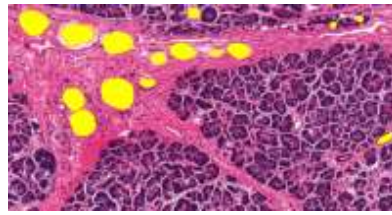


Figure 22 (b)- Tagged image

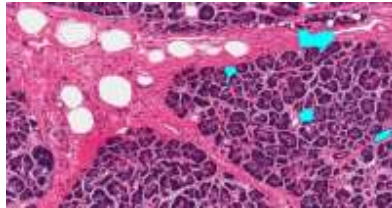


Figure 22 (c)- Adiposoft

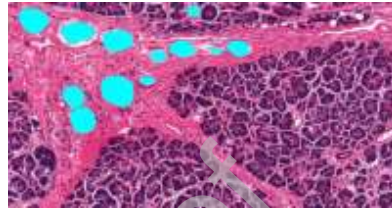


Figure 22 (d)- Fatquant

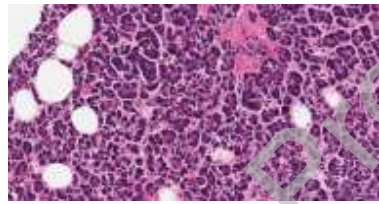


Figure 23 (a)- Raw image

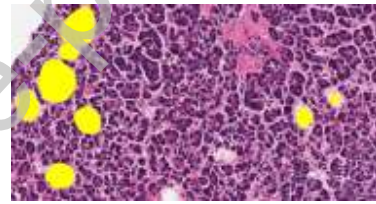


Figure 23 (b)- Tagged image

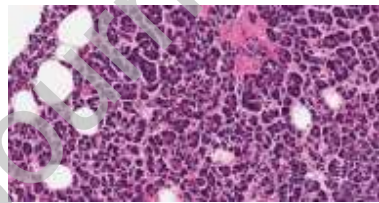


Figure 23 (c)- Adiposoft

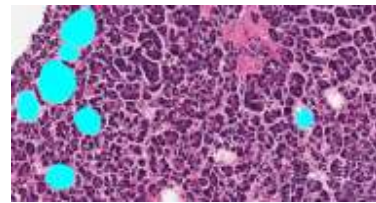
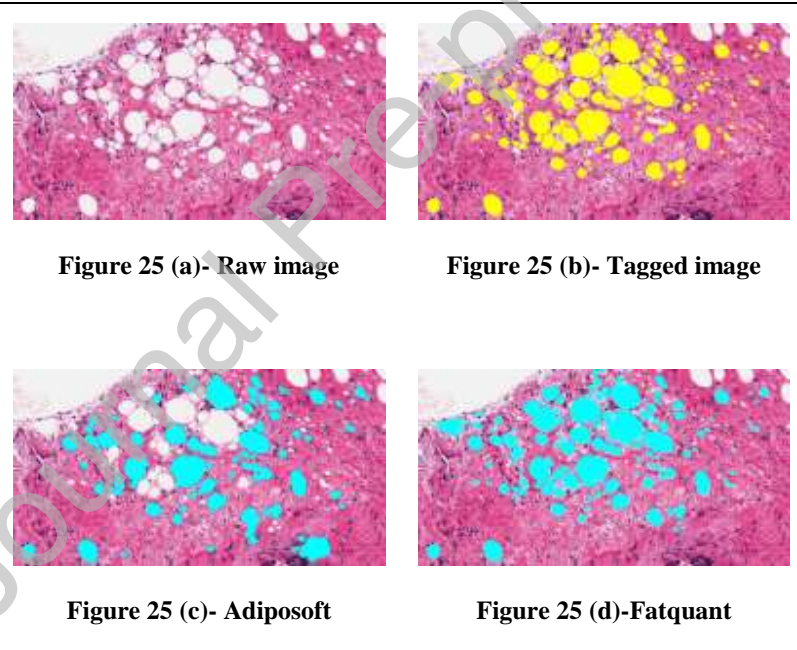
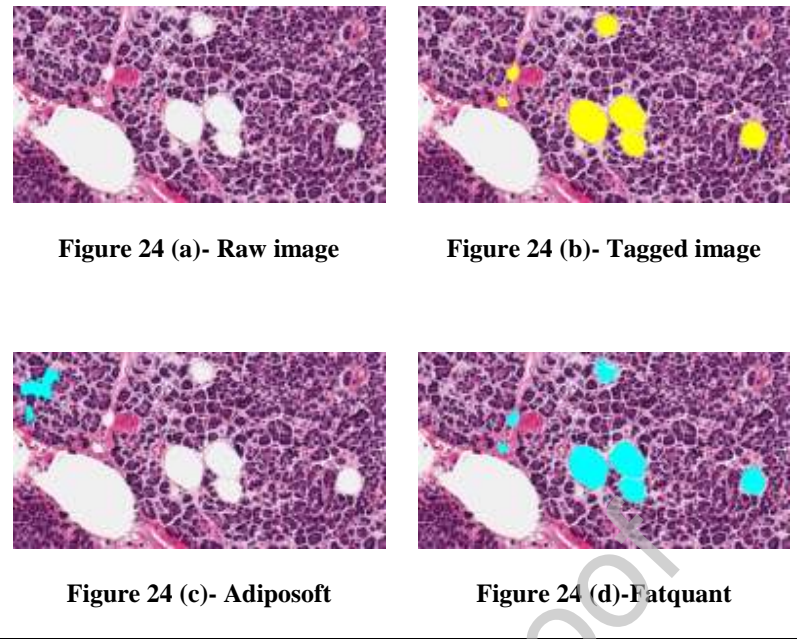


Figure 23 (d)- Fatquant



Figures 16 – 25 are from 7 unique subject IDs and 3 random images with different fat percentage/neighboring cells. Image (a) are the raw sample histology images of pancreas; (b) are manual fat tagged images; (c) and (d) are outputs from Adiposoft and Fatquant respectively.

Table 1: Manually chosen parameters used by Adiposoft and Fatquant for optimal output in

Figures 16 – 25 sample pancreatic histology images

Figures	Samples	GTEx subject ID	Adiposoft		Fatquant		
			Min. diameter	Max. diameter	Threshold value	Min. diameter	Max. diameter
16	1	11DXZ-0826	20	135	230	27	130
17	2	1122O-0726	25	500	228	25	500
18	3	1117F-1726	22	500	228	29	500
19	4	1117F-1726	20	180	233	24	150
20	5	117YW-0926	20	280	231	24	250
21	6	117YW-0926	22	280	229	27	250
22	7	13PVQ-2026	22	250	226	29	200
23	8	13FHP-1926	40	300	228	40	250
24	9	13FHP-1926	22	210	228	22	190
25	10	11WQC-0926	14	200	231	14	200

Table 2: Calculation of accuracy on pancreas sample histology images in figures 16 – 25

Samples	Tool	TP (in pixels)	FP (in pixels)	FN (in pixels)	IoU
1	Fatquant	110429	20387	871	0.839
	Adiposoft	0	0	111300	0
2	Fatquant	81041	5597	468	0.930
	Adiposoft	46769	60732	34740	0.329
3	Fatquant	591209	49758	11326	0.906

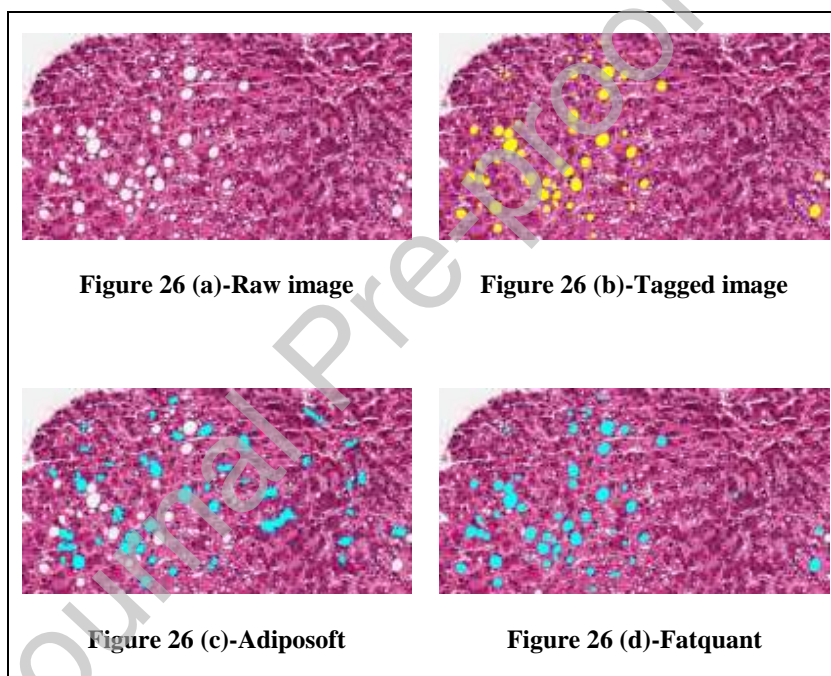
	Adiposoft	571949	14599	30586	0.927
4	Fatquant	72793	13053	5455	0.797
	Adiposoft	0	3245	78248	0
5	Fatquant	899818	35098	8576	0.954
	Adiposoft	851499	2285	56895	0.935
6	Fatquant	195357	5405	3457	0.957
	Adiposoft	0	6674	198814	0
7	Fatquant	81834	146	5830	0.932
	Adiposoft	2065	14838	85599	0.020
8	Fatquant	91516	1106	2498	0.962
	Adiposoft	0	0	94014	0
9	Fatquant	75566	2377	1319	0.953
	Adiposoft	0	15468	196051	0
10	Fatquant	219816	13070	8254	0.912
	Adiposoft	134342	64357	93728	0.459

The parameter values of Adiposoft and Fatquant (mentioned in Table 1) are not exactly same because the default threshold or edge detection values used by Adiposoft are unknown to users. Hence for analysis, values have been chosen as per optimal output. Pixels being part of fat in (b) are marked with yellow color whereas pixels identified as fat in (c) and (d) are marked with cyan color.

From the outputs it can be noted that Adiposoft only shows decent output when adipocytes cover maximum area of a sample image (e.g. Figures 18 and 20). In a heterogeneous sample image this tool can tag many non-fat areas as valid fats (e.g. Figures 21, 22 and 24). Moreover, it can even

fail to identify presence of any fat in an image (e.g. Figures 16 and 23). Hence, Fatquant performs better in the scenarios shown. The tagging of fat cells shown in (b) images are done by the authors as no ground truth data of the slides was available.

Additionally, ten liver histology images were also analyzed and are available in ‘Liver samples’ directory of our GitHub repository. IoU of range 0.675 to 0.937 and 0 to 0.497 was identified using Fatquant and Adiposoft respectively, indicating wide applicability of our tool. The parameters used and accuracy on each image are shown in Tables 3 and 4. The outputs are shown below.



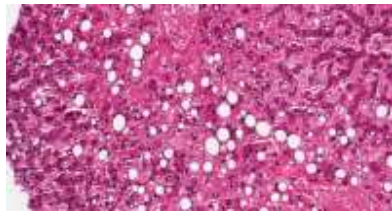


Figure 27 (a)- Raw image

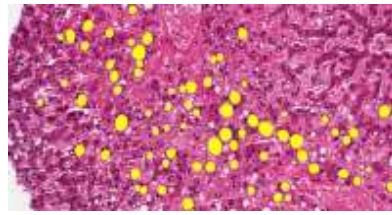


Figure 27 (b)- Tagged image

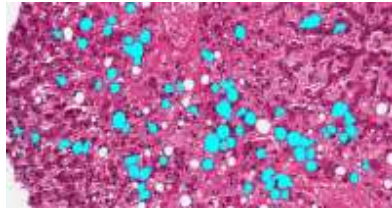


Figure 27 (c)- Adiposoft

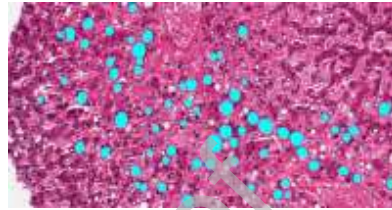


Figure 27 (d)- Fatquant



Figure 28 (a)- Raw image

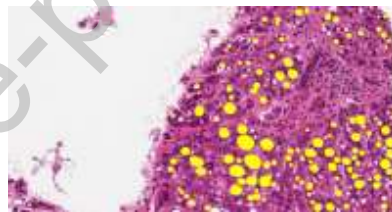


Figure 28 (b)- Tagged image



Figure 28 (c)- Adiposoft

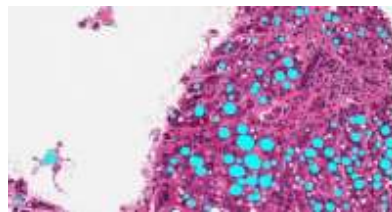


Figure 28 (d)- Fatquant

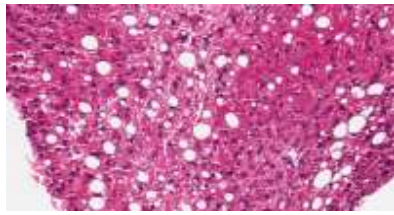


Figure 29 (a)- Raw image

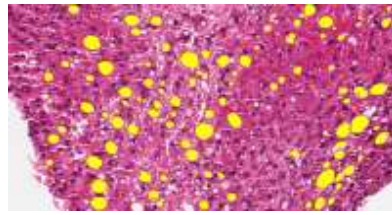


Figure 29 (b)- Tagged image

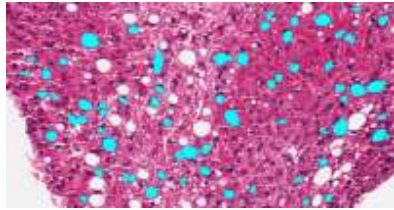


Figure 29 (c)- Adiposoft

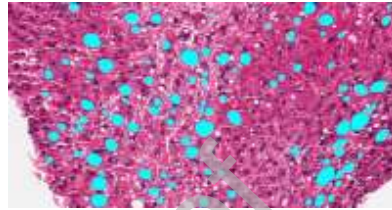


Figure 29 (d)-Fatquant

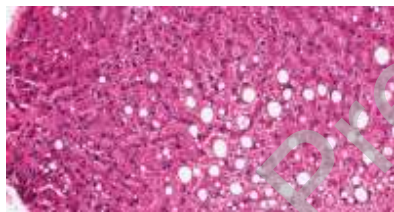


Figure 30 (a)- Raw image

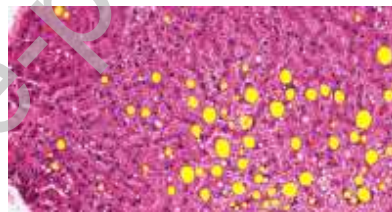


Figure 30 (b)- Tagged image

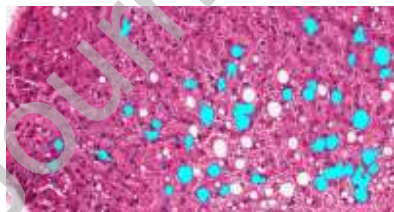


Figure 30 (c)- Adiposoft

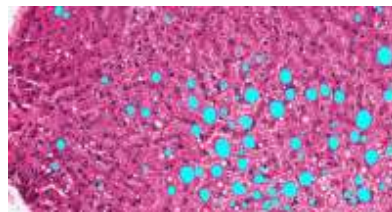


Figure 30 (d)-Fatquant

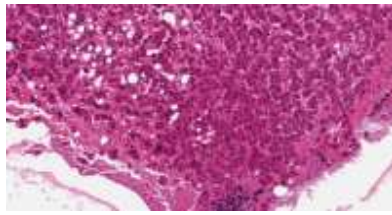


Figure 31 (a)- Raw image

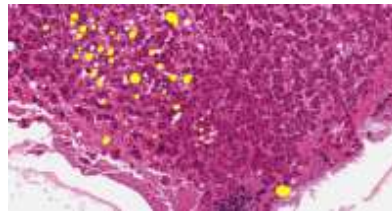


Figure 31 (b)- Tagged image

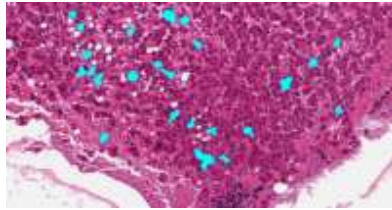


Figure 31 (c)- Adiposoft

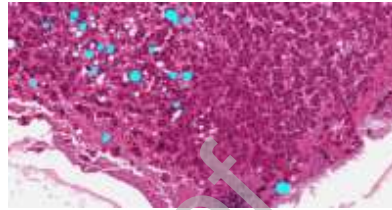


Figure 31 (d)-Fatquant



Figure 32 (a)- Raw image

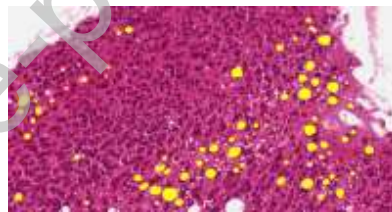


Figure 32 (b)- Tagged image

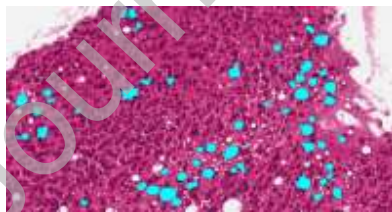


Figure 32 (c)- Adiposoft

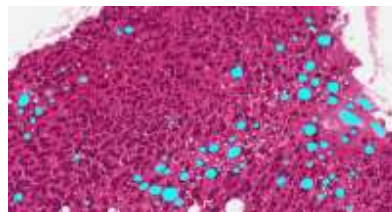


Figure 32 (d)-Fatquant

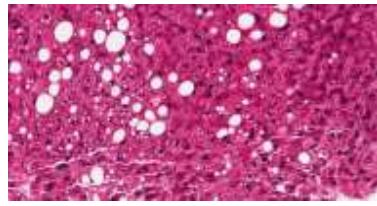


Figure 33 (a)- Raw image

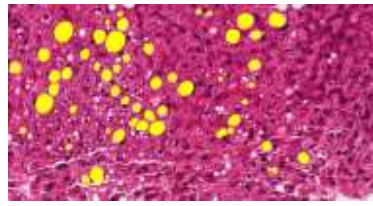


Figure 33 (b)- Tagged image

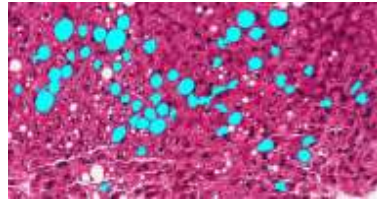


Figure 33 (c)- Adiposoft

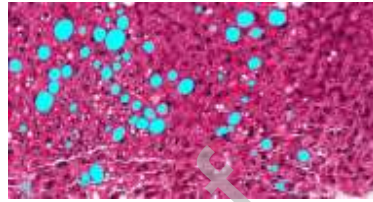


Figure 33 (d)-Fatquant

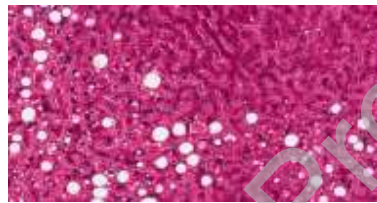


Figure 34 (a)- Raw image

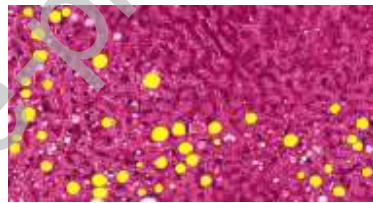


Figure 34 (b)- Tagged image

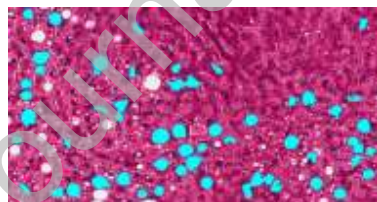


Figure 34 (c)- Adiposoft

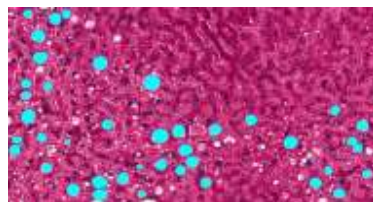


Figure 34 (d)-Fatquant

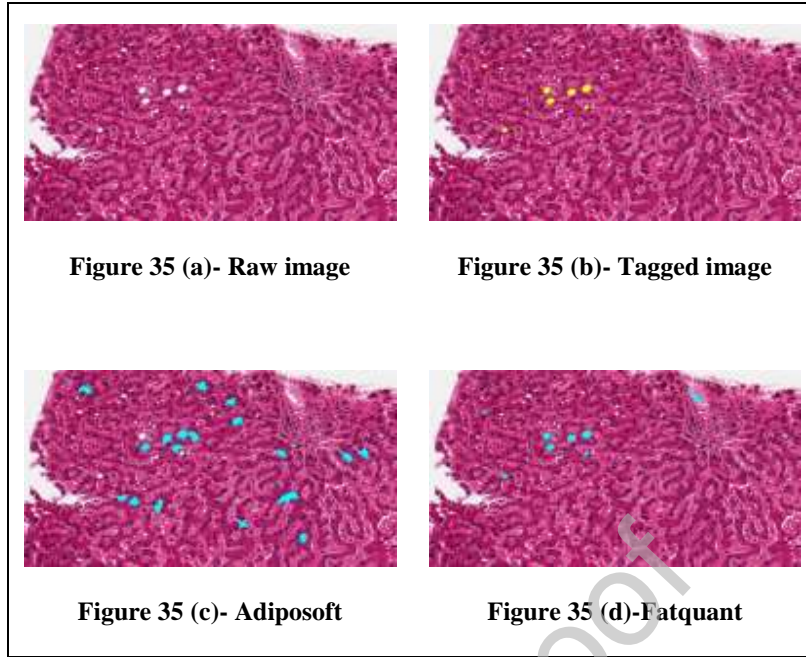


Figure 26-35 include six unique subject IDs and rest random images with different fat percentage/neighboring cells. Raw (a), manual tagged (b), Adiposoft (c) and Fatquant (d) outputs of each sample liver histology images.

Table 3: Manually chosen parameters used by Adiposoft and Fatquant for optimal output in figure 26-35 sample liver histology images

Figures	Samples	GTEx subject ID	Adiposoft		Fatquant		
			Min. diameter	Max. diameter	Threshold value	Min. diameter	Max. diameter
26	1	11EMC-0326	14	100	227	18	90
27	2	11EMC-0326	18	42	226	22	100
28	3	11EQ9-0526	14	100	226	16	100
29	4	11GS4-0926	14	100	226	18	100
30	5	11GS4-0926	19	100	226	18	100

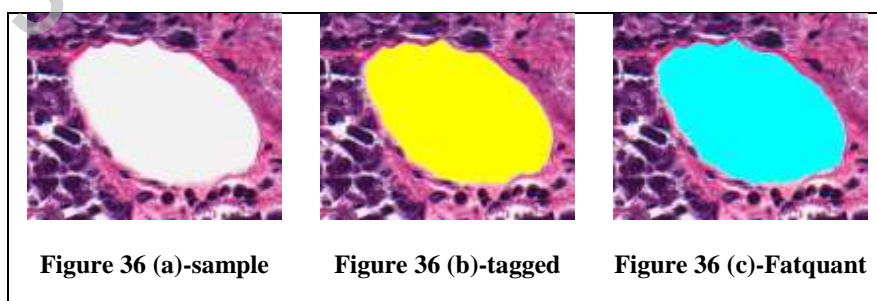
31	6	11GSO-0826	16	95	227	22	90
32	7	11GSO-0826	13	34	226	14	90
33	8	11NV4-1326	19	90	226	18	100
34	9	11NV4-1326	19	90	226	22	90
35	10	11O72-0926	18	22	223	16	80

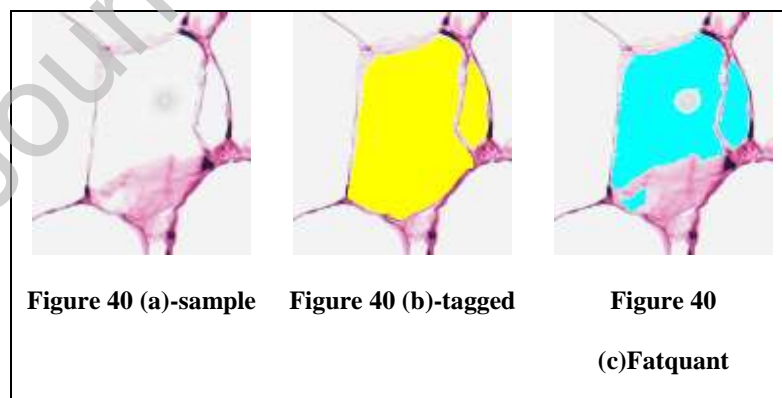
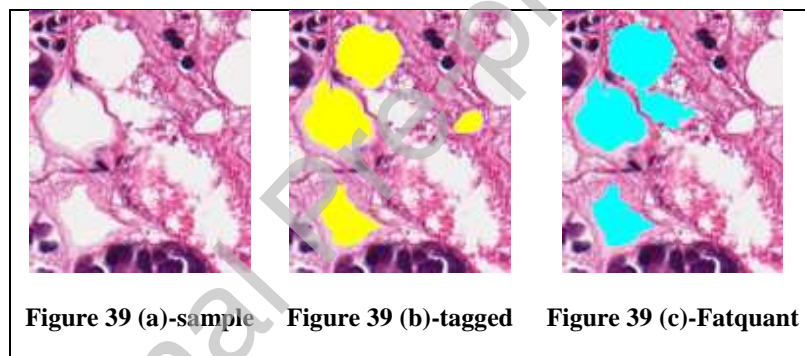
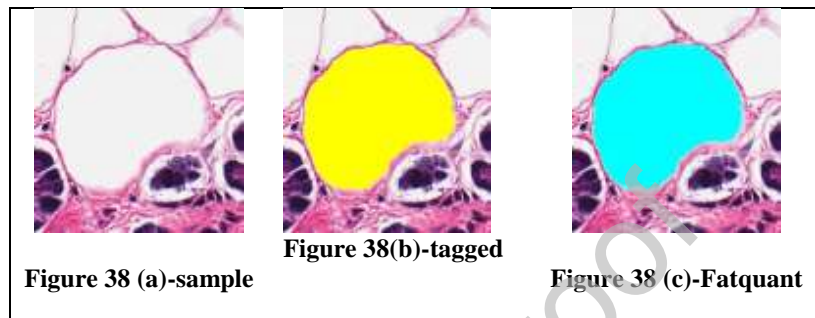
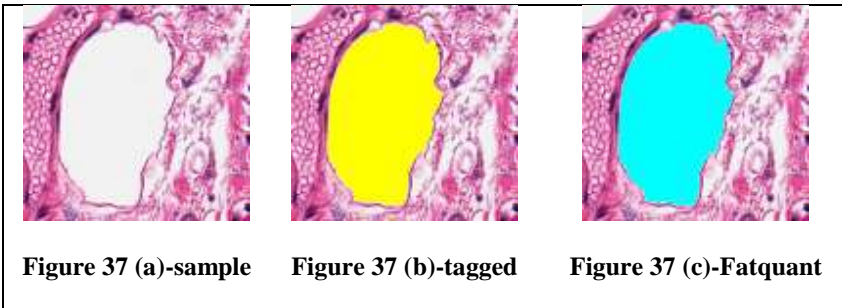
Table 4: Calculation of accuracy on liver sample histology images in figures 26 – 35

Samples	Tool	TP (in pixels)	FP (in pixels)	FN (in pixels)	IoU
1	Fatquant	53527	3496	1291	0.918
	Adiposoft	23480	50420	31338	0.223
2	Fatquant	82426	2970	2563	0.937
	Adiposoft	54410	59912	30579	0.375
3	Fatquant	82873	5357	3868	0.900
	Adiposoft	0	1698	86741	0
4	Fatquant	109075	9588	5708	0.877
	Adiposoft	47210	45192	67573	0.295
5	Fatquant	83690	5350	2165	0.918
	Adiposoft	37072	58771	48783	0.256
6	Fatquant	13776	1162	5700	0.668
	Adiposoft	9563	23556	9913	0.222
7	Fatquant	49570	8228	1259	0.839
	Adiposoft	38614	30704	12215	0.474
8	Fatquant	93577	6793	1685	0.917

	Adiposoft	80351	66466	14911	0.497
9	Fatquant	75279	2126	4524	0.919
	Adiposoft	58962	71464	20841	0.390
10	Fatquant	4368	1466	15	0.747
	Adiposoft	2639	22297	1744	0.099

Image (a) in Figures 36 – 41 are the sample images; (b) represents manual tagged and (c) represents Fatquant tagged areas of images in (a). As per the data mentioned in Table 5 it can be noted that fat cells available in Figures 36 – 38 are easy to get tagged by this tool, hence IoU value increases. Whereas the tool does not detect fat cell properly when cell boundaries are not clearly delineated as seen in Figures 39 – 41, hence IoU value decreases. So validity analysis performed on sample images with many fat cells similar to that of Figures 36 – 38 will likely show higher accuracy. Manual tagged data shown in (b) images are not collected from any laboratory website and are rather created by the authors. So, there can be some variation in ground truth data of these images created by any other source. In the GitHub repository mentioned, sample images used in Figures 16 – 25 and Figures 36 – 41 are available in ‘Pancreas_samples’ and ‘Small_pancreas_samples’ directories respectively.





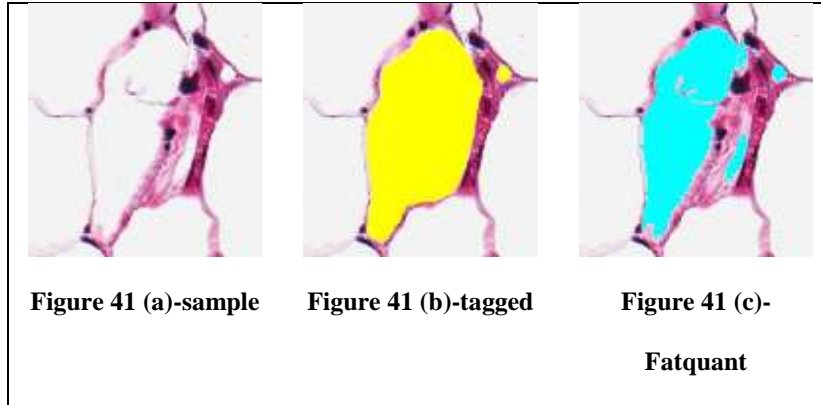
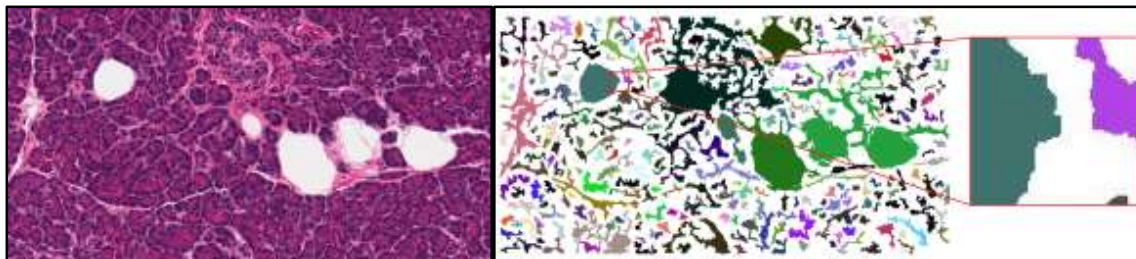


Table 5: Accuracy of annotation by Fatquant in Figures 36 – 41

Figures	GTEx subject ID	Threshold value	Min. diameter	Max. diameter	TP (in pixels)	FP (in pixels)	FN (in pixels)	IoU
36	117YW- 0926	229	20	180	17640	118	102	0.987
37	1117F- 1726	227	30	200	19321	269	15	0.985
38	117YW- 0926	228	20	200	26095	115	334	0.983
39	111VG- 0926	227	35	100	5692	1312	326	0.776
40	117YW- 0926	236	15	220	18178	499	7042	0.707
41	117YW- 0926	231	10	200	16441	484	7623	0.670

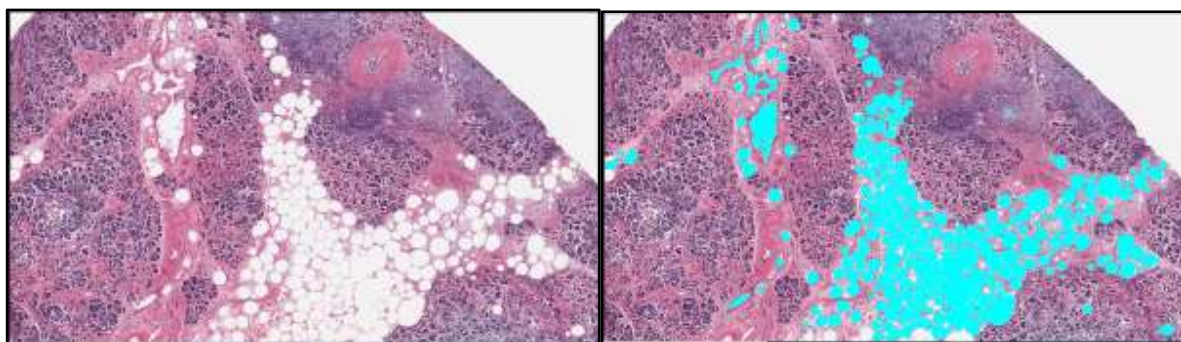
We also analyzed the H and E images using AdipoCount in which the detected fat areas are colored and non-fat areas are white but both of which contain noise.



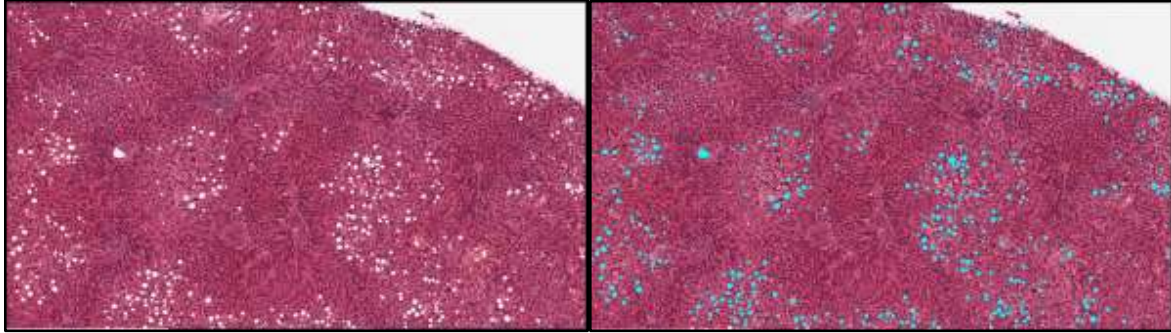
Sample H and E raw (left) image and AdipoCount (right) output image

AdipoCount output (right) where each detected fat lacks a clear outline and is tagged with a range of similar colors is due to the image getting stored in lossy JPEG file format. Whereas the image output in Adiposoft can be stored in lossless file formats (TIF and PNG) because of which the issues mentioned above does not occur here and detected fat areas can be properly segmented and used for comparison with manually annotated image.

Finally, we have demonstrated Fatquant output with images of higher resolution indicating applicability of our tool with whole slide images. However, due to limitations of RAM (12GB) on our system, we would need access to system of higher RAM (32GB or 64GB) to further validate tool with whole slide images.



Pancreas histology raw image (left) and Fatquant (right) output labeled in cyan color for a pancreas 8K resolution (7680 x 4320 pixels) image (GTEx-1117F-1726) with parameters: Binary Threshold value: 230, Minimum fat diameter: 55, Maximum fat diameter: 340.



Liver histology raw image (left) and Fatquant (right) output labeled in cyan color for a liver 8K resolution (7680 x 4320 pixels) image (GTEx-11NV4-1326) with parameters: Binary Threshold value: 226, Minimum fat diameter: 19, Maximum fat diameter: 150.

4. Discussion

Pancreatic tumor segmentation [19, 25-27] and QuPath based fat analysis [28] are deep learning approaches which require large training datasets [28]. Liver steatosis detection methods are currently either not open source software [14] or deep learning based methods [29]. The Fatquant tool helps to identify fat cells in heterogeneous histological tissue sections which would complement work of pathologists. The growing importance of fat cells in pathophysiological functions and with availability of whole slide digitized images, Fatquant can save significant amount of time for image analysis. Based on our analysis, Adiposoft shows decent output when adipocytes cover maximum area of a sample image and label non-fat areas as valid fat cells in heterogeneous tissue sample images. In absence of ground truth images, we manually annotated fat cells in different pancreas and liver sample images from a public portal and then compared the output accuracy using IoU with Fatquant tool annotated results. We observed an intersection over union of 0.797 to 0.962 and 0.675 to 0.937 for pancreas and liver, respectively exhibiting high degree of similarity between manual versus Fatquant annotation. Moreover, we have shown

utility of our Fatquant tool for 8K resolution images and works well for both homogenous fat tissue and varied heterogenous tissues like pancreas, liver suggesting diverse applicability of the tool for analysis of fat cells.

Acknowledgments

There is no potential conflicts of interest and no funding to declare. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The employers of the authors (R.R.N & A.R) have no role whatsoever towards this submitted work.

References

1. Yu, T. Y., & Wang, C. Y. (2017). Impact of non-alcoholic fatty pancreas disease on glucose metabolism. *Journal of diabetes investigation*, 8(6), 735–747. <https://doi.org/10.1111/jdi.12665>
2. Kochar, I. S., & Jain, R. (2021). Pancreas transplant in type 1 diabetes mellitus: the emerging role of islet cell transplant. *Annals of pediatric endocrinology & metabolism*, 26(2), 86–91. <https://doi.org/10.6065/apem.2142012.006>
3. Verma, A. R., & Papalois, V. (2011). Evaluating steatosis in pancreatic transplant. *Experimental and clinical transplantation : official journal of the Middle East Society for Organ Transplantation*, 9(3), 159–164.
4. Dholakia, S., Sharples, E. J., Ploeg, R. J., & Friend, P. J. (2017). Significance of steatosis in pancreatic transplantation. *Transplantation reviews (Orlando, Fla.)*, 31(4), 225–231. <https://doi.org/10.1016/j.trre.2017.08.001>

5. Paul, J., & Shihaz, A. (2020). PANCREATIC STEATOSIS: A NEW DIAGNOSIS AND THERAPEUTIC CHALLENGE IN GASTROENTEROLOGY. *Arquivos de gastroenterologia*, 57(2), 216–220. <https://doi.org/10.1590/s0004-2803.202000000-27>
6. Lesmana, C.R.A., Pakasi, L.S., Inggriani, S. *et al.* Prevalence of Non-Alcoholic Fatty Pancreas Disease (NAFPD) and its risk factors among adult medical check-up patients in a private hospital: a large cross sectional study. *BMC Gastroenterol* **15**, 174 (2015). <https://doi.org/10.1186/s12876-015-0404-1>
7. Tariq, H., Nayudu, S., Akella, S., Glandt, M., & Chilimuri, S. (2016). Non-Alcoholic Fatty Pancreatic Disease: A Review of Literature. *Gastroenterology research*, 9(6), 87–91. <https://doi.org/10.14740/gr731w>
8. Fukui, H., Hori, M., Fukuda, Y., Onishi, H., Nakamoto, A., Ota, T., Ogawa, K., Ninomiya, K., Tatsumi, M., Osuga, K., Yamada, D., Eguchi, H., Miyoshi, E., & Tomiyama, N. (2019). Evaluation of fatty pancreas by proton density fat fraction using 3-T magnetic resonance imaging and its association with pancreatic cancer. *European journal of radiology*, 118, 25–31. <https://doi.org/10.1016/j.ejrad.2019.06.024>
9. Virostko J. (2020). Quantitative Magnetic Resonance Imaging of the Pancreas of Individuals With Diabetes. *Frontiers in endocrinology*, 11, 592349. <https://doi.org/10.3389/fendo.2020.592349>
10. Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B.

(2009). Histopathological image analysis: a review. *IEEE Rev Biomed Eng*, 2, 147-171. doi: 10.1109/RBME.2009.2034865.

11. Galarraga, M., Campion, J., Munoz-Barrutia, A., Boque, N., Moreno, H., Martinez, J. A., . . . Ortiz-de-Solorzano, C. (2012). Adiposoft: automated software for the analysis of white adipose tissue cellularity in histological sections. *J Lipid Res*, 53(12), 2791-2796. doi: 10.1194/jlr.D023788.

12. Zhi, X., Wang, J., Lu, P., Jia, J., Shen, H. B., & Ning, G. (2018). AdipoCount: A New Software for Automatic Adipocyte Counting. *Front Physiol*, 9, 85. doi: 10.3389/fphys.2018.00085.

13. Glastonbury, C. A., Pulit, S. L., Honecker, J., Censin, J. C., Laber, S., Yaghootkar, H., . . . Lindgren, C. M. (2020). Machine Learning based histology phenotyping to investigate the epidemiologic and genetic basis of adipocyte morphology and cardiometabolic traits. *PLoS Comput Biol*, 16(8), e1008044. doi: 10.1371/journal.pcbi.1008044.

14. Nativ, N. I., Chen, A. I., Yarmush, G., Henry, S. D., Lefkowitz, J. H., Klein, K. M., Maguire, T. J., Schloss, R., Guarrera, J. V., Berthiaume, F., & Yarmush, M. L. (2014). Automated image analysis method for detecting and quantifying macrovesicular steatosis in hematoxylin and eosin-stained histology images of human livers. *Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society*, 20(2), 228–236. <https://doi.org/10.1002/lt.23782>

15. Catanzaro, R., Cuffari, B., Italia, A., & Marotta, F. (2016). Exploring the metabolic

syndrome: Nonalcoholic fatty pancreas disease. *World journal of gastroenterology*, 22(34), 7660–7675. <https://doi.org/10.3748/wjg.v22.i34.7660>

16. Apaolaza, P. S., Petropoulou, P. I., & Rodriguez-Calvo, T. (2021). Whole-Slide Image Analysis of Human Pancreas Samples to Elucidate the Immunopathogenesis of Type 1 Diabetes Using the QuPath Software. *Frontiers in molecular biosciences*, 8, 689799. <https://doi.org/10.3389/fmolb.2021.689799>

17. Floros, X., Fuchs, T. J., Rechsteiner, M. P., Spinas, G., Moch, H., & Buhmann, J. M. (2009). Graph-based pancreatic islet segmentation for early type 2 diabetes mellitus on histopathological tissue. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 12(Pt 2), 633–640. https://doi.org/10.1007/978-3-642-04271-3_77

18. Rechsteiner, M. P., Floros, X., Boehm, B. O., Marselli, L., Marchetti, P., Stoffel, M., Moch, H., & Spinas, G. A. (2014). Automated assessment of β -cell area and density per islet and patient using TMEM27 and BACE2 immunofluorescence staining in human pancreatic β -cells. *PloS one*, 9(6), e98932. <https://doi.org/10.1371/journal.pone.0098932>

19. Huang, Y., Liu, C., Eisses, J. F., Husain, S. Z., & Rohde, G. K. (2016). A supervised learning framework for pancreatic islet segmentation with multi-scale color-texture features and rolling guidance filters. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 89(10), 893–902. <https://doi.org/10.1002/cyto.a.22929>

20. Consortium, G. TEx. (2015). Human genomics. The Genotype-Tissue Expression (GTEx)

pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235), 648-660. doi: 10.1126/science.1262110.

21. Badea, L., & Stanescu, E. (2020). Identifying transcriptomic correlates of histology using deep learning. *PLoS One*, 15(11), e0242858. doi: 10.1371/journal.pone.0242858.

22. Dite, P., Blaho, M., Bojkova, M., Jabandziev, P., & Kunovsky, L. (2020). Nonalcoholic Fatty Pancreas Disease: Clinical Consequences. *Digestive diseases (Basel, Switzerland)*, 38(2), 143–149. <https://doi.org/10.1159/000505366>

23. Sequeira, I. R., Yip, W. C., Lu, L., Jiang, Y., Murphy, R., Plank, L. D., Cooper, G., Peters, C. N., Lu, J., Hollingsworth, K. G., & Poppitt, S. D. (2022). Pancreas Fat, an Early Marker of Metabolic Risk? A Magnetic Resonance Study of Chinese and Caucasian Women: TOFI_Asia Study. *Frontiers in physiology*, 13, 819606. <https://doi.org/10.3389/fphys.2022.819606>

24. Rugivarodom, M., Geeratrakool, T., Pausawasdi, N., Charatcharoenwitthaya, P. (2022). Fatty Pancreas: Linking Pancreas Pathophysiology to Nonalcoholic Fatty Liver Disease. *J Clin Transl Hepatol*. 1-11. doi: 10.14218/JCTH.2022.00085.

25. Vu, Q. D., Graham, S., Kurc, T., To, M., Shaban, M., Qaiser, T., Koohbanani, N. A., Khurram, S. A., Kalpathy-Cramer, J., Zhao, T., Gupta, R., Kwak, J. T., Rajpoot, N., Saltz, J., & Farahani, K. (2019). Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Frontiers in bioengineering and biotechnology*, 7, 53. <https://doi.org/10.3389/fbioe.2019.00053>

26. Yang, C., Xiang, D., Bian, Y., Lu, J., Jiang, H., Zheng, J. (2021). Gland segmentation in pancreas histopathology images based on selective multi-scale attention. *Proc. SPIE 11596, Medical Imaging 2021: Image Processing*, 115962M (15 February 2021); <https://doi.org/10.1117/12.2581039>
27. Huang, M., Huang, C., Yuan, J., & Kong, D. (2021). A Semiautomated Deep Learning Approach for Pancreas Segmentation. *Journal of healthcare engineering*, 2021, 3284493. <https://doi.org/10.1155/2021/3284493>
28. Palomäki, V. A., Koivukangas, V., Meriläinen, S., Lehenkari, P., & Karttunen, T. J. (2022). A Straightforward Method for Adipocyte Size and Count Analysis Using Open-source Software QuPath. *Adipocyte*, 11(1), 99–107. <https://doi.org/10.1080/21623945.2022.2027610>
29. Roy, M., Wang, F., Vo, H., Teng, D., Teodoro, G., Farris, A. B., Castillo-Leon, E., Vos, M. B., & Kong, J. (2020). Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Laboratory investigation; a journal of technical methods and pathology*, 100(10), 1367–1383. <https://doi.org/10.1038/s41374-020-0463-y>

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof