Quantitative Analysis of Lipid Droplets using k means image processing

Ankur Abhijeet
Department of Mathematical Sciences
NISER, Jatni,752050
1911026
ankur.abhijeet@niser.ac.in

Summit Bikram Nayak
Department of Mathematical Sciences
NISER, Jatni,752050
2011167
summit.nayak@niser.ac.in

Abstract

This report provides an overview of the use of K-means clustering in the quantification of lipid area in cell pictures. Lipids play essential roles in cellular processes, and accurate quantification of lipid area is critical for studying lipid metabolism and related diseases. The K-means clustering algorithm is an unsupervised machine learning technique that can group similar objects into clusters based on their features. The algorithm is applied to cell pictures by preprocessing the image, extracting features, and clustering the features into k groups, where k represents the number of lipid areas in the image. The algorithm's advantages include scalability, flexibility, and unsupervised learning. Limitations include cluster initialization, sensitivity to outliers, and determining the optimal number of clusters. Overall, K-means clustering is a useful tool for quantitatively estimating lipid area in cell pictures.

1 Introduction

The study of lipid droplets in cells has become an area of intense research interest due to their crucial role in energy storage, metabolism, and signaling. Accurately quantifying lipid droplets within cells is essential to understanding their function and potential implications in diseases such as obesity and diabetes. Traditional methods of quantifying lipid droplets have been laborious and time-consuming, leading researchers to turn to machine learning algorithms as a faster and more efficient alternative. In this paper, we explore the use of machine learning in estimating quantitative analysis of lipid droplets in a given cell. We present our findings on the accuracy and efficiency of machine learning algorithms in lipid droplet quantification and compare them to traditional methods. Our study provides a promising avenue for future research into the use of machine learning in lipid droplet analysis and its potential applications in disease diagnosis and treatment.

2 Related Works

1.

https://www.researchgate.net/publication/256291891_A_novel_automated_image_analysis_method_for_accurate_adipocyte_quantification

The paper titled "A novel automated image analysis method for accurate adipocyte quantification" presents a new image analysis approach for the automated quantification of adipocytes in histological sections. The authors use a combination of color deconvolution and machine learning techniques to analyze images of tissue stained with hematoxylin and eosin, a commonly used stain in histology.

Preprint. Under review.

The approach is based on the segmentation of individual adipocytes using a random forest classifier, followed by the quantification of the total area of adipocytes in each histological section. The algorithm is validated against manual counting by two independent observers, demonstrating a high degree of accuracy and reproducibility.

2.

`https://www.jlr.org/action/showPdf?pii=S0022-2275%2820%2931064-6`

The paper titled "Deep Learning for Quantitative Analysis of Lipid Droplets in Cells" published in the Journal of Lipid Research, presents an investigation into the use of deep learning techniques for the quantitative analysis of lipid droplets in cells. The paper is well-structured, providing a clear introduction to the importance of studying lipid droplets in cells, as well as a comprehensive review of related work.

The authors use a deep learning-based approach for lipid droplet quantification, which involves training a convolutional neural network (CNN) on a dataset of fluorescent images of cells stained with Nile Red. The performance of the CNN is evaluated in terms of lipid droplet count and size measurements, and compared against traditional methods such as manual counting and threshold-based analysis.

3 Baseline Algorithm

K-means is a widely used unsupervised machine learning algorithm that is commonly used for clustering and segmentation tasks. Each data point is assigned to a cluster based on its distance from the centroid of that cluster. K-means can be applied to various types of data, including images, text, and numerical data, and has been used in a wide range of applications such as image segmentation, document clustering, and customer segmentation. While K-means is a simple and efficient algorithm, its performance can be sensitive to the initial choice of centroids, and it may not always converge to the optimal solution. Despite these limitations, K-means remains a popular baseline algorithm for unsupervised machine learning tasks, including the quantification of lipid droplets in stained cell pictures.

4 Experimental Details

4.1 Github

The code for the project is in the provided link:

`https://github.com/summit-b/machine_learning.git`

4.2 Steps undertaken

There are several steps procedure to achieve our goal.

4.2.1 Dataset

Variety of lipid stained cell dataset are avaliable online but for our use we borrowed some samples from Prof. Chandan Goswami's Lab. the provided dataset is of a high quality Oil Red O lipid stained cell photograph with very low noise.

4.2.2 Image Preprocessing

The first step is to preprocess the image by removing any noise or artifacts that may interfere with the analysis.
We skipped this step as we already had a photo with very low noise, in case of noise we can use different photi filters to obtain a low noise picture.

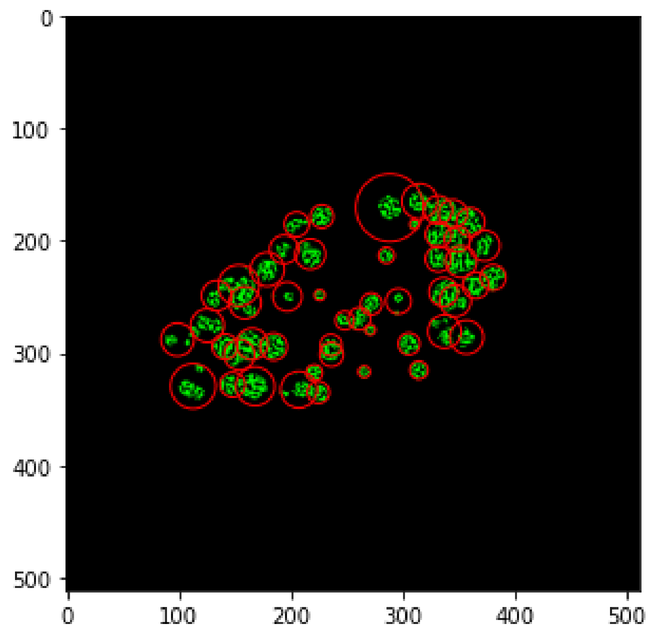

Figure 1: figure

4.2.3 Image Segmentation

The next step is to segment the image into different regions, where each region represents a potential lipid droplet.
The provided photo is to be trimmed and color adjusted manually if needed.

4.2.4 Feature Extraction

The third step involves extracting features from each segmented region such as differentiating on the basis of color.
The stained part in the picture provided was found to be having a 255 in green pixels for the lipid portions.

4.2.5 K-means Model

Using K Means model with manual input of k and then using elbow method to get a optimum k value.
The photo was converted, stored and processed as a numpy array for using k means.

4.2.6 Clustering

now the model can be used to cluster the segmented regions into different groups, where each group represents a lipid droplet.
The model clusters the lipid droplets in form of drops to identify the specific segments as drops of liquid.

4.2.7 Post-Processing

The quantification of lipid area can be done by calculating the area of each cluster centroid. The final step involves post-processing the results to obtain the final lipid droplet quantification by comparing the lipid area obtained to the area obtained by the cell boundary picture.

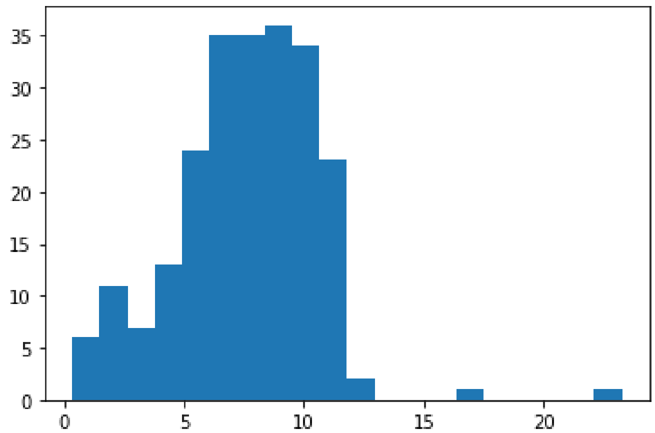

Figure 2: figure

4.3 Limitations faced

1.At first we tried to differentiate on the basis of pixel value after converting the image to greyscale, then we faced problems in selecting the optimum pixel and the shifted to green scaling in the RGB segment, helping us to identify the segments with stained lipids.

2.The next hurdle faced was to get the optimum k value. The elbow method comparing the total area along with number of centroids provides us with a suitable k value. But in this case we were not able to obtain a sharp drop in area for an optimum k value. Thus, we chose it at a point that seems feet manually.

3.The most important problem faced is determining the radius of individual centroids as some small green fragments are considered to be part of the circle increasing the expected area.

5 Plans

1.Finding out a way to workout the problem with elbow method or find a way to work around it.

2.In cases such as in the "figure" shown , we can try to work with dividing the big empty circles multiple smaller ones.

3.Working out a solution to find the cell boundary in the same picture.

6 Unventured scopes

One potential avenue for further exploration is to combine K-means clustering with other machine learning algorithms to improve accuracy and precision in lipid area identification. For example, support vector machines (SVMs) can be used to classify pixels as lipid or non-lipid, which can then be used to guide the K-means clustering process. This hybrid approach has shown promising results in identifying lipid droplets in various cell types.

Combining K-means clustering with fluorescent labeling of lipids could be explored to enhance the specificity and sensitivity of the technique.

7 References

[1] Osman, Omnia Selway, Joanne Kepczynska, Malgorzata Stocker, Claire O'Dowd, Jacqueline Cawthorne, Michael Arch, Jonathan Jassim, Sabah Langlands, Kenneth. (2013). A novel automated image analysis method for accurate adipocyte quantification. Adipocyte. 2. 160-4. 10.4161/adip.24652.

[2] Digital image analysis approach for lipid droplet size quantitation of Oil Red O-stained cultured cells Author links open overlay panelManuel J. Deutsch, Sonja C. Schriever, Adelbert A. Roscher, Regina Ensenauer Research Center, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-Universität München, 80337 Munich, Germany